

77634

10/4 04 + 02

Access DB#

10/4 115

SEARCH REQUEST FORM

Scientific and Technical Information Center

78201

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: Oct 4, 2002
 Art Unit: 1652 Phone Number 305-7270 Serial Number: 09/720,583
 Mail Box and Bldg/Room Location: 10 D.O.G Results Format Preferred (circle): PAPER DISK E-MAIL
10.D01

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Propionibacterium Vector
 Inventors (please provide full names): Pieter Pauwels et al

Earliest Priority Filing Date: 6/25/99

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO: 1 DNA
 2 AA
 3 AA

Thank you so much in advance.

M. Walicka

Point of Contact:
 Toby Port
 Technical Info. Specialist
 CM1 6A04
 703-308-3634

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: _____	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>2</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>10/4</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>10/7</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>12</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>12</u>	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)


```

/note="unnamed protein product"
/transl_table=1
/codon_start=1
/protein_id="CAC07380.1"
/db_xref="GI:9995030"
/translation="MDSFETLFEPSWLPKRPPLAASEKSGAVRHVTRORALELPYEAN
PLVMSVLTITRDSDADMDADLAGLSPSPSVSNRRVTGGHIYALKNPVCLTDAR
RRPENTLARVEOGICDVLGGDPAASGHRITLSTPATITMGSPADALYELRALAHITD
EIALHPAGNRRNRVTRSTVGRNVLTDTRMAVRAVRRMSGCPVAEMHHYAEHITH
LNTNTITADEFATPGLINELNKLHSRSISKRWNRNFTPEFRANOKAISLGASGGK
EGHGKGIGASASRRNRHTRROOFLGLS
1181. 1438
CDS
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07381.1"
/db_xref="GI:9995031"
/translation="MTTRERLPKRGYSITAAAKLGVSESTVYKRWISPEDEFYARVA
ARRHRIKELRREGSGSMALIAVEGVSQVTHYALNKNRTDA"
705 a 1156 c 1075 g 619 t
BASE COUNT
ORIGIN

```

[illegible][illegible]

QY 1801 ccgagacgagtagaacccgctccggttgggtggtgacactcgtgactcaaggtcgtt 1860
|||
Db 1801 ccggagccacgagtagaacccgctccggttgggtggtgacactcgtgactcaaggtcgtt 1860
|||
QY 1861 ttgggggttgcagtagaagaccctcgaaatgccccggtcactgaagcttgaatgtggcgagc 1920
|||
Db 1861 ttgggggttgcagtagaagaccctcgaaatgccccggtcactgaagcttgaatgtggcgagc 1920
|||
QY 1921 tgcgtctcgaagaacagagtttcttaacggtctcaggtcgaagcagagcgagcagac 1980
|||
Db 1921 tgcgtctcgaagaacagagtttcttaacggtctcaggtcgaagcagacgagcagac 1980
|||
QY 1981 ggcagagctgaagcagacggttgcctcagcagcagctcaacggttgcgagcgttgaag 2040
|||
Db 1981 ggcagagctgaagcagacggttgcctcagcagcagctcaacggttgcgagcgttgaag 2040
|||
QY 2041 ccggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2100
|||
Db 2041 ccggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2100
|||
QY 2101 ggtgagcagcgtcgtgtagcgttgcagtagcgtctcgtgagcaggtgtctgcagagtcg 2160
|||
Db 2101 ggtgagcagcgtcgtgtagcgttgcagtagcgtctcgtgagcaggtgtctgcagagtcg 2160
|||
QY 2161 aacaggttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
|||
Db 2161 aacaggttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
|||
QY 2221 agttcgtcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2280
|||
Db 2221 agttcgtcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2280
|||
QY 2281 agttcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2340
|||
Db 2281 agttcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2340
|||
QY 2341 actgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400
|||
Db 2341 actgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400
|||
QY 2401 agtgcgtcgttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2460
|||
Db 2401 agtgcgtcgttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2460
|||
QY 2461 agtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2520
|||
Db 2461 agtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2520
|||
QY 2521 gctcgtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2580
|||
Db 2521 gctcgtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2580
|||
QY 2581 ccgaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2640
|||
Db 2581 ccgaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2640
|||
QY 2641 gtagagctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2700
|||
Db 2641 gtagagctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2700
|||
QY 2701 ggcgagc 2760
|||
Db 2701 ggcgagc 2760
|||
QY 2761 cgatcggtgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2820
|||
Db 2761 cgatcggtgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2820
|||
QY 2821 aagtttctgatacgtcaatgcatgtagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2880
|||
Db 2821 aagtttctgatacgtcaatgcatgtagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2880
|||

QY 2881 caacgagcagatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2940
|||
Db 2881 caacgagcagatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2940
|||
QY 2941 ccccaagc 3000
|||
Db 2941 ccccaagc 3000
|||
QY 3001 ggttcgagc 3060
|||
Db 3001 ggttcgagc 3060
|||
QY 3061 gaacggttcagc 3120
|||
Db 3061 gaacggttcagc 3120
|||
QY 3121 agactcagc 3180
|||
Db 3121 agactcagc 3180
|||
QY 3181 cggcagc 3240
|||
Db 3181 cggcagc 3240
|||
QY 3241 tcatgagc 3300
|||
Db 3241 tcatgagc 3300
|||
QY 3301 acggtcagc 3360
|||
Db 3301 acggtcagc 3360
|||
QY 3361 aacgaagc 3420
|||
Db 3361 aacgaagc 3420
|||
QY 3421 atctcagc 3480
|||
Db 3421 atctcagc 3480
|||
QY 3481 gatgagc 3540
|||
Db 3481 gatgagc 3540
|||
QY 3541 cagagc 3555
|||
Db 3541 cagagc 3555
|||

RESULT 2
AX008203 3555 bp DNA linear PAT 06-SEP-2000
LOCUS
DEFINITION Sequence 1 from Patent WO967356.
ACCESSION AX008203
VERSION AX008203.1 GI:9995826
KEYWORDS
SOURCE
ORGANISM
Proionibacterium freudenreichii.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Proionibacteriaceae; Proionibacteriaceae;
Proionibacterium.
REFERENCE
AUTHORS Jore,J.P., Luiten,R.G., Pouwels,P.H. and Van Luit]K.N.
TITLE Proionibacterium vector
JOURNAL Patent: WO 967356-A 1 29-DEC-1999;
GIST BROCHADES BV (NL); JORE JOHANNES PETRUS MARIA (NL); LUIJK
NICOLE VAN (NL); LUITEN RUDOLF GUSBERTUS MARIE (NL); POWELS
PIETER HENDRIK (NL)
FEATURES
location/Qualifiers
1..3555
/organism="Proionibacterium freudenreichii"
/db_xref="taxon:1744"
/note="TMG16545"

CDS
273..1184
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07396.1"
/db_xref="GI:9995827"
/translation="MDSFETLPPEEWLPKPKPLASAKSAYRHVTRORALELYPEAN
PLVMSLVITDDASDADMAADLAGLPSPSYSVMNVTGHTVALKNPVCLTDAR
RPIINLAEVQGLCDVLGSDASYGHRITKNPJSTAHAILMGFADLYELRALAHLD
EIHAIPEANPRNRYSTYGRNVLTFTDRMAAYAVHSMGSPVAEMHWYFEHII
LNETIADPEFATPLGLINELKLSISIRWWRNPTPTFARAKAISRASKGK
EGHKGIASGASRAHTRQFLKGLS"
1181..1438
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07397.1"
/db_xref="GI:9995828"
/translation="MTTERLPRNGYSIAAAKKIGVSESTYKRWTSREPEFVARVA
ARRARIRELRSBQSMRAIAAEVGSVGVHYALNKRIDA"
BASE COUNT 705 a 1156 c 1075 g 619 t
ORIGIN

Query Match 100.0%; Score 3555; DB 6; Length 3555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtgcaccctgacagccgagagcaagtcagcgaaagtcacagctgcgcgaggaacta 60
DB 1 gtgcaccctgacagccgagagcaagtcagcgaaagtcacagctgcgcgaggaacta 60
QY 61 gccgcaatgccgcaaacagcccccagccatcccttgagacagctgcgcgaggaact 120
DB 61 gccgcaatgccgcaaacagcccccagccatcccttgagacagctgcgcgaggaact 120
QY 121 cggggagatgtttgacagggatgtgaaagagatgccttctgcaatgagtcacag 180
DB 121 cggggagatgtttgacagggatgtgaaagagatgccttctgcaatgagtcacag 180
QY 181 ggttaactaactgatatgggtcttcgtgcgccaacttgaaacagccgaggaatggaac 240
DB 181 ggttaactaactgatatgggtcttcgtgcgccaacttgaaacagccgaggaatggaac 240
QY 241 gctgaacgtgactgcagctgcttcaactgcatgta tgaatctgctcagaagctgttccctg 300
DB 241 gctgaacgtgactgcagctgcttcaactgcatgta tgaatctgctcagaagctgttccctg 300
QY 301 agagctgagctgcacagcaagccgctgctgcagcgcgaagaagcttgaggcgtaaccgagc 360
DB 301 agagctgagctgcacagcaagccgctgctgcagcgcgaagaagcttgaggcgtaaccgagc 360
QY 361 tgaactcgcagagagggcgctgagctgcttcaactgcaagaggaacccgttgatcatgagt 420
DB 361 tgaactcgcagagagggcgctgagctgcttcaactgcaagaggaacccgttgatcatgagt 420
QY 421 ccttgatcatalcaacagatcgagatgcttcgagatgctgagctgcgcgagacactgcgtggc 480
DB 421 ccttgatcatalcaacagatcgagatgcttcgagatgctgagctgcgcgagacactgcgtggc 480
QY 481 tgcctcaccgctcctcagcttcaatgaaacgtgtcaagcaacacgcgaacacatgctctatg 540
DB 481 tgcctcaccgctcctcagcttcaatgaaacgtgtcaagcaacacgcgaacacatgctctatg 540
QY 541 ccttgaagaacccctgtgtgtctgacagatgcgcgcgagagagcctataaactgctg 600
DB 541 ccttgaagaacccctgtgtgtgtctgacagatgcgcgcgagagagcctataaactgctg 600
QY 601 cccgcgttcgagcaaggcctcatgcaagcttctgcgcgcgagatgcatccttaacggcaccgga 660
DB 601 cccgcgttcgagcaaggcctcatgcaagcttctgcgcgcgagatgcatccttaacggcaccgga 660
QY 661 tcacaaagaacccgctcagcaacgcgcacatgcaacccctctgaggcccccgcagacgagctct 720

DB 661 TCACAAAGAACCCGCTCAGCAGCCGACCCATCCGACCCCTTGGGGCCCCGCGAGAGCCCTCT 720
QY 721 acaagctgcgcgcctctgcagacacacccctgcagagagttcgaagctgcgcgagagcgagg 780
DB 721 ACAAGCTGCGCCCTCTCGACACACCCCTCAGCAGATCCAGCAGATGCGCGAGGCGAGGGA 780
QY 781 accgcgttcgcagacgtcaccgga tcaagagctgcgcgcgcaacgctgcctgttcgaacca 840
DB 781 ACCCGGTCGCAACGTACACCCGATCAACGGTGGCGCAACGTCACCCCTTGTGACACCA 840
QY 841 cccgcatgtggaatacagcagggcgctgcgcgacatcctgaggcgccggttcgcgaatagg 900
DB 841 CCGGCAATGTGGGATACCGGGCGCTCGGACATCTGGGGCGCGCGCTCGCGGAATGGG 900
QY 901 agaacaccgatttcgagacatacactactgaagagagaatcatcgcgcgaagaattg 960
DB 901 AGCAACCCGATTTCGAGCAGATCCACTACTGAACGAGATCATGCGCGAAGATTCG 960
QY 961 ccaagggcccccctgcgcttgaaacgaactaagcactatctcgatccatctccgattgg 1020
DB 961 CCAAGGCCCCCTCGCGCTTGAAGCACTTAAGCACTTAATCGATCATTTCCGATGGG 1020
QY 1021 tctgagcaacttcaccccccgaaccccttcgcgcagcgcgaagaagcatcagctcgtg 1080
DB 1021 TCTGGCGCAACTTCACCCCGCAACCTTCGCGCAGCCGAGAAAGCATCAGCTCCGTG 1080
QY 1081 gaacatccaaaggcgcgcaaaagagcgcgccacaaaggcgccatgagtgagcgtacac 1140
DB 1081 GAGATTCCAAAAGCGGGGAAAGAGGCGGCCACAAAGCGGCAATGGCAGTGGCGCATCAC 1140
QY 1141 ggcgcgcccaataccgctcaacagcttcttgagggtctctcatatgacacacagtgtaagctc 1200
DB 1141 GGGCGGCCATACCCGTCACACAGTTCCTGAGGGGTCTCTATACCAACAGTAAAGTGT 1200
QY 1201 ccccgcgaacgcttaacagcatcgcgctgctgcgaaagaagctcgtgtctcgaatccac 1260
DB 1201 CCCCCGAAAGGCTTACAGCATCCCGCTCGCGGAAAGCTCGGTCTCCAGTCCAC 1260
QY 1261 cgtcaagcgttggaacttcgcagcagcgcgagaggttcgtgcccgcgcttgccgacgca 1320
DB 1261 GGTCAACGGGTGACATTCCTCCAGCCACGCGAGAGTTCGTGGCCCCGCTGGCCACGCCA 1320
QY 1321 cgcgcgattcgtgagctccgctcgcgagaggttcagagcatgctgctgcagatgtgcgaggt 1380
DB 1321 CCGCGGATTCGTGAGCTCCGCTCGAGGGGTCAAGAGATCGTGCATGCCATGCCGAGGT 1380
QY 1381 cgggggttccgtgaggcagctgcaatcagcgcgctgaaagaatcgaaatgagcatgacc 1440
DB 1381 CGGGGTTTCCGTGGGACCGTGCACACCGCGTGAACCAAGATCGAACCGCATGACC 1440
QY 1441 gtaacgcgcaagatggaatttcttgatcgtgcagccgcttggaactaagttcagctg 1500
DB 1441 GTAAACCGGACATGACATTTTCTTGTGATCGTGACACCGTTGGCATCTACGCTGCG 1500
QY 1501 ggttgacagctgcgcgcacagcttctatcctgcgcgcatgtgtgctacagacgaatgggg 1560
DB 1501 GGTTCGACAGATGCGCGCACGTTCTTATCTCGCGCATGTGGGTAAACGCCAATGGGG 1560
QY 1561 gcatcagcaaggaagcttgaaacccggttggaaggtttactcaagggggagatgccagttc 1620
DB 1561 GCATCAGCAACGAGCATTTGAACCCGGTGGCAAGTGTATCTAGGGGGGAAATCCACAGTC 1620
QY 1621 tgcgcgcctcgaatgagcgttatggagatgctgcatgagcggcccaacgctcaactatc 1680
DB 1621 TCGCGGCTCGGATTGACGGTATGGCAGTCTGTCATGCGGCCCCACCGTCAAACTATTTC 1680
QY 1681 aggtatcagtgaaacacctatgcaacccctcgtgtacacgttctcgtgtgcatacagctg 1740
DB 1681 AGGTATCAGTGAACACCTCATGAGCAACCCCTCTGTGACACGTTCTGTGGCATCAGCTG 1740
QY 1741 ctgtgcgtgcggcgctgagcgttcttaagctgcgcgcgcaagaaatcaagacgttaggctg 1800

Db 1741 CTGTCGTCGGGGCGTGAAGCTTTTACGCTTCGGCGCAGGAATCAGAGCTTGAAGCTTG 1800
 Qy 1801 ccgagacgacgagtaagaccgctccggttggtgtgtccactgctgtaactcaaggtctgtt 1860
 Db 1801 CCGGAGCGACGGGTAGACCCGCTCGGTTGGGTGTGCACTGCTGACACTCAAGAGTCTGTTT 1860
 Qy 1861 ttgggtgtcagatagagacactgaaatgagcccggtcaatgagctgagttagtgtgcagc 1920
 Db 1861 TTGGGGGTCTCAATAGACACTGGAATGCGCCCGGTCTATGACCTGATGAGTGGGCGAAC 1920
 Qy 1921 tgcgtctgagaagagattttaaaggcttcaagctcagagagaaggcgcgaactcgaac 1980
 Db 1921 TGCCTCTGAGAACAGATTTTATACGGCGCTCAGAGTCGAGAGAGGCGCCAGATCGAAC 1980
 Qy 1981 ggcagagctgagagacacacgcggtgtgtctcagcgcaagctcacaagggtttgccaagttaag 2040
 Db 1981 GGCAGAGCTGAGGACACACGGGCTGTCTCAGCGACGCTCAGAGGTTGGCCAGCTTGAGG 2040
 Qy 2041 ccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2100
 Db 2041 CCGGCGAGCGACACACACGACACTGCGCACCCGTTGAAAGCGCGGCTCCGCGACGGCGGT 2100
 Qy 2101 ggtgacgagctgt 2160
 Db 2101 GGTGGACGCTGT 2160
 Qy 2161 aacagctgctgacacagctgt 2220
 Db 2161 AACAGTTCCTGACACAGT 2220
 Qy 2221 agtgcgtgagagactgt 2280
 Db 2221 AGT 2280
 Qy 2281 agtgcagagctgt 2340
 Db 2281 AGTTCAGAGCTGT 2340
 Qy 2341 actgagacgagcagagagatgt 2400
 Db 2341 ACTGAGACCGGCGGAGAGATGT 2400
 Qy 2401 agtgcgt 2460
 Db 2401 AGT 2460
 Qy 2461 aggtacattccgagcagctgt 2520
 Db 2461 AGGTACATTCCGGGCGCTGT 2520
 Qy 2521 gctcgtgatactgt 2580
 Db 2521 GCTCTGT 2580
 Qy 2581 ccgagcttcaatgt 2640
 Db 2581 CCGAGCTTCATGTCCAGCGCAGCAGCTGAAGAGACCAAGCGAGTATACCTGCTAAC 2640
 Qy 2641 gtagagctgt 2700
 Db 2641 GTATAGCTTGT 2700
 Qy 2701 ggcgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2760
 Db 2701 GCGGCGAAGCGGCTACCGGCTTGGGCGACAGCCCAATTAATCAACCGCTCCGGTGTGAAG 2760
 Qy 2761 cgatgt 2820
 Db 2761 CGATGT 2820
 Qy 2821 aagttttttgt 2880
 Db 2821 AAGTTTTCGTATAGCTAATGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2880

Qy 2881 caacgagcagatctcaggt 2940
 Db 2881 CAACGAGCAGATTCAGGATGTGGGCGAGAGCGCGAGGCGGATACAGACTCCCAACT 2940
 Qy 2941 ccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000
 Db 2941 CCCCAAGCCAGCGCGGCGAGCGCGCGCGCTAGAGAGCGGTCCGGGCGCGTACCCGT 3000
 Qy 3001 ggtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3060
 Db 3001 GGTCTCGACGGGCGCCAGCTGTGCGCTCTCACAGAACGACAGCAAGCGCGAGGCGATTCAC 3060
 Qy 3061 gaacgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3120
 Db 3061 GAACGCTTACAGCGGATTCGAGCGCAGTCCAGAGTGTGACAGCGGTTGCTGACCTCC 3120
 Qy 3121 acgactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3180
 Db 3121 ACGACTCAGCAGCAGCAAGCACTACCAAGAGACCGGCTCGAGCAGACAGCGCGTGTCTACG 3180
 Qy 3181 cggcagc 3240
 Db 3181 CGGCAGCCAGCTTCTCAACTCCGGCGCACTGACGACGACGACGACGACGACGACGACGACG 3240
 Qy 3241 tcatgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3300
 Db 3241 TCATGATCGGAACCGACCGACGAGCGCGCTACTGCAATCTGTGCTGATCTACGACG 3300
 Qy 3301 acggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3360
 Db 3301 ACGGCTACAGATGATCATCCACGCAATGAAAGCCGCGACCCCAATTCGACAGCTCT 3360
 Qy 3361 aaccaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3420
 Db 3361 AACCAAGAAAGAACTGT 3420
 Qy 3421 atcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3480
 Db 3421 ATCTTCAGAGT 3480
 Qy 3481 gatgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3540
 Db 3481 GATGAAGCCGCTGT 3540
 Qy 3541 cagc 3555
 Db 3541 CAGCAGCTGCGGCC 3555

RESULT 3
 AF291751
 LOCUS AF291751 3555 bp DNA circular BCT 01-FEB-2001
 DEFINITION Propionibacterium freudenreichii plasmid p545, complete sequence.
 ACCESSION AF291751
 VERSION AF291751.1 GI:10953186
 KEYWORDS
 SOURCE
 ORGANISM
 Propionibacterium freudenreichii.
 Propionibacterium freudenreichii Bacteria; Firmicutes; Actinobacteridae; Actinomycetales; Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
 REFERENCE
 1 (bases 1 to 3555)
 Jore,J.P., van Luijk,N., Luiten,R.G., van der Werf,M.J. and Poneels,P.H.
 Efficient transformation system for Propionibacterium freudenreichii based on a novel vector
 JOURNAL Appl. Environ. Microbiol. 67 (2), 498-503 (2001)
 MEDLINE
 2 (bases 1 to 3555)
 Jore,J.P.M. and van Luijk,N.
 Direct Submission
 JOURNAL Submitted (31-JUL-2000) AMCT, TNO Nutrition and Food Research,

Utrechtseweg 48, Zeist 3704 HE, The Netherlands
Location/Qualifiers
1. .3555
/organism="Proionbacterium freudenreichii"
/plasmid="p545"
/strain="LMG16545"
/db_xref="taxon:1744"
28. .485
/note="putative replication protein"
/codon_start=1
/transl_table=11
/product="rep2"
/protein_id="AAC25295.1"
/db_xref="GI:10953187"
/translation="MTTRELPRNGYSIAAAAKLGVSSESTVSRMTSEPREFVARVA
ARHARIRELSESGSRAIAEVEGVSQVHYALANKNRIDA"
join(2875..3555,1..231)
/note="putative replication protein"
/codon_start=1
/transl_table=11
/product="rep1"
/protein_id="AAC25294.1"
/db_xref="GI:10953187"
/translation="MDSFETLPESWLPKPLASAKSGARVHTROALEPYIAN
PLVMOSLVITDRDASDADNADLAGLPSRSTVSMKRYTTGHTVTLKMPVCLTDAR
RRPILNLAHQGLCDVLDGDAYGHRITKPNLSTAHALWGPADALYELRALAHLD
EIALPBAQPRRNVTSTVGRNVTLEDTRMVAAYRAVRSMGSPVAEMHTEFHH
LNMETIADFEATPGLNELKHLRSISISWVRNPTPEFRAKQAIISDGSKSGK
EGSHKGGISGASRRARHTROQPLEGLS"
BASE COUNT 705 a 1156 c 1075 g 619 t
ORIGIN

Query Match 73.2%; Score 2602; DB 1; Length 3555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 954 gaattgcacaaagccctcgcttgaaacgaacttaagcaattatcgcattcc 1013
DB 1 GAATTCGCCACAGGCCCTCGCTGACGAACCTATCATCATTCATTC 60

QY 1014 cgaatggctcggcgcaactcaaccccgaaacctccgagcgacggaagaagc 1073
DB 61 CGATGGCTCGGCGCAACTTACCCCGAATCTTCGCGACCCGAAAGCGATCAGC 120

QY 1074 ctccgtgagcatcacaagcgcgcaagaagcgcgccacaagcgcgcatgca 1133
DB 121 CTCGCTGAGCATCCAAAGGCGCAAGAGCGGCCCAACAAGCGCATTCAGTGC 180

QY 1134 gcatcagcgcgcccataccgctcaacagttcttgagggtctctcaatgaaca 1193
DB 181 GCATCAGCGCGCGCCCATACCCGCAACAGTTCTTGGAGGGTCTCTCATGCACACAGTGC 240

QY 1194 aagctcccccgaacgagctaacagcaltgcgcgtgctggaaaaacgtcg 1253
DB 241 AAGCTCTCCCCGGAAGGCTACAGCATCGCGCTGCTGGAAAAGCTGCTCTCCG 300

QY 1254 agtcacagctcaagcggtggaacttcgagccagcgagagttcggtggcc 1313
DB 301 AGTCACACGTCGAACGGGTGAGCTTCGAGCCAGCGAGAGATTGCTGGCCGCTTCCG 360

QY 1314 caagcagcgcggaattcgtgaagctcgcgtclogaagggtcagaagcalt 1373
DB 361 CACGCCACGCGCGGATTCGTGACGCTCGCGTCCGAGGGTCAAGCATGCTGCATTGCTCG 420

QY 1374 ccgaggtcgggggttcctcgtgggacacgltgactacgcgtgaaacaaga 1433
DB 421 CCGAGGTGCGGGGTTTCGTGGGCAACGTCGACATACGCGCTGAACAAATGAACTGACG 480

QY 1434 catgacgtaacgcgcagcagatgagcattctctgacgtcgacacgcttgca 1493
DB 481 CATGACGTAACGCGCGACATGAGCATTTTCTTGATCGTGCACCGCTTGCACTAGCTT 540

QY 1494 cgcgtcgattgcaacagltgcgcgacacgtctctatccctcggccattggt 1553
DB 541 CGCGTCGGGTTCACAGTGGCGCCGACGTTTATCTTGCGGCATGTGGTACACCA 600

QY 1554 atggggggatcagcaacgagcgtttgaaccccggttggcgaagtgttactca 1613
DB 601 ATGGGGGGCATCACAACGAGCGTTGAACCCGGGGGCAAGTGTATCATCAGGGGACATG 660

QY 1614 cccagctcgcgcgtcgatgacgttatgacgtctgtacgtctgacgtcgcc 1673
DB 661 CCCAGTCTGGCGGCGCTCGGATGTGACGGTATGACAGTGTGATCGGCCGCCACCGTCAAA 720

QY 1674 ctcaatgaagltacgtgaaacccctcaatgcaacccctcgtgacacgtt 1733
DB 721 CTCATTCAGTATCATGAGAAACCCCTATGGACCCCTCGTGACAGTTCCTGCTGCA 780

QY 1734 tcagctcgtgtggttggcggtgagcgtttctcaagctcgtgagcgaggaata 1793
DB 781 TCAGCTCTGTGCGTGGCGGTGAGCGTTTCTACGCTGGCGGCAAGAAATCAGACCTT 840

QY 1794 gaagctcgcggaagcagcgtgaaacccgtccggttggttggtgacacgtc 1853
DB 841 GAGGCTCCCGAGCGGAGCGTATGAGACCCGCTCGGTTGGGTGGTCCACCTGCTGCACTAAG 900

QY 1854 gtccgtttgggtgtgcaagatgaaacctcgaatgcccgcgtcaatgacgt 1913
DB 901 GTCGTTTGGGGGTGTGATGATGAACTCAATCCGACCCGCTCATGACGCTGAGTTAGTG 960

QY 1914 ggcgagctggtctctgaaacgagtttctcgggtcaggttcgagcagcagcg 1973
DB 961 GCGCAGCTGCGCTCTGAAACGAGTTTTCAGGGGTATGAGTTCGAGCAGCAGCGCGCAGC 1020

QY 1974 atcgaacgcaagcgttgagcacaacgcgtgtcctcaagcgacgtcacaag 2033
DB 1021 ATCGAAGCGCAGCGGTGAGGCAACACGCGGTGTCTACACGCACTCAACGCGTTGGCAG 1080

QY 2034 ctgaaagcgcgcgcagcagaacgacacacgtgacccgttgaagaagccgc 2093
DB 1081 CTTGAGCGCCGCGCGACGACGACGACCACTGACCCGTTGAAAGCGCGGCTCCGCA 1140

QY 2094 cggcggtggtgagcggtcgtgtagcgttcaggtcgtcgtggtgaggtgt 2153
DB 1141 CGCGGTTGGTGGACGCTCGGTACGGTCAAGATCGCTTGCGCTGACGAGTGTGTGG 1200

QY 2154 caatgcaacagttgctcgcacgaatgacgaagaagcagacgtcgtggtgt 2213
DB 1201 CAGTGCAGAACGTTGCTCGACGACGACGACGACGACGACGACGACGACGACGACGACG 1260

QY 2214 ctgcgtcagttcgtcgaagcagtcgaggtctgctgcgtccaaagcgaat 2273
DB 1261 CTCGCTCAGTTCTGTCGAGGACCTGCGGGGTCTGCTGCTCAGCCGCTCGCGGCG 1320

QY 2274 caaggtcagttcgaagcgttgcgaacgacgagcgccctcggtgacagc 2333
DB 1321 CAAAGTCAATTCGAAGCTGTGCAACGACACGCGCTCGGCTGACAGCTGAGCTTGAA 1380

QY 2334 ctgtgcaactggaacgcgcgcgaagatgacgttgcgaagttcgtgaaatg 2393
DB 1381 CTGTGCAACTGGAACGCGCGCGAAGATGACAGTTCGAGTGTGCGCAAGCGCAC 1440

QY 2394 gtcgaagagtgctgtctcgtgaagcgcgaagaatgagcagttcgtgca 2453
DB 1441 GTCAAAAGTGTCTGTCGTAAGCGCGCAAAATGCGAGTGTGCGGATTCGGAATCGG 1500

QY 2454 cctgtccaggttaacttcgcggcgcttgatgaaacgctccgctgaagaag 2513
DB 1501 CCTGCTCCAGTACATTCCTGGGCGCGCTTGATGAACGCTTCGCGTGAAGGCGCACCGTTCT 1560

QY 2514 cggcgcgacctgtgacgtctcaactgtgaagcgtccctcgtatggtttctga 2573
DB 1561 CGCGCGCGCCTCGTCACTGTGCACTGTGCACTGCTCTCTGATGTTCTTGACGCTACCG 1620

QY 2574 gagaacacgaagcgttcatgcccagcgacgacgtgaagaagcaagcgaagt 2633

```

Db 1621 GAAACACGACGACGTTATGCCCCAGGACGACCTAAGAGCAAGCCAGTTAGCCGT 1680
Oy 2634 gctaacgctatagctgtctccgtcgcctctgtagggcaacacactgtgcagcaagtggcg 2693
Db 1681 GCTAACCGTATAGCTTGCTCCGCTCTGAGGGCAACCACTGGCCACAGATGGGCG 1740
Oy 2694 gcaagccgcgcgaagcgcctacacgggtttgagcagaccataalcaacgcctccggt 2753
Db 1741 GCAGCCCGGCGGACGCGCTACCGGGTTTGGCCACAGCCCAATCAACGCGCTCCGCT 1800
Oy 2754 gttgaagcgatcggtgtcaagatgtctatgcttctacccttcaggglttcgtatac 2813
Db 1801 GTTGAAGCGATCTGTGTCAAGATTGCTATAGCTTGTGCTTCAAGGGTTTCGTATAC 1860
Oy 2814 acaaatcaagtttttcgtatacgtctaatgccaatgtagagcactactacacgaagc 2873
Db 1861 ACAATCAAGTTTTCGTATACGCTAATGCTAATGAGTGAAGCACTACTGACGCAAGC 1920
Oy 2874 ccgtcaccaacgaagcagatcgaagcaltgagcagaagcgaagccggaatacgaacctgc 2933
Db 1921 CCGTCACCAAGAGAGATTCAGAGCATGGGCAAGAGCGCGGAGCGCGATACGACCTGC 1980
Oy 2934 ccaaaatccccaaagcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2993
Db 1981 CCAAACTCCCAAGCAAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Oy 2994 taccggtgctctcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3053
Db 2041 TACCCGCTGCTCGAGCGCGCGCAACCGTTGCCCTCTCAAGANAGCAAGCCGAGG 2100
Oy 3054 gcatcagaacgcgtltcagaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3113
Db 2101 GCATCAAGAACCGTTGAGAGCGGATCCGAGCCGAGTCCAGAGTGGACACAGGGTTGCCCT 2160
Oy 3114 gacctcagaactcagcaagcaagcaactacaaagcaagcgcgcgcgcgcgcgcgcgcgcgc 3173
Db 2161 GACCTTCACAGACTCAGACAGCAAGCACTACCAAGAGACCGGCTCGAGACGCGCGCTG 2220
Oy 3174 ctctaagcgagcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3233
Db 2221 CTCTACGCGGCAACCAACGTTCTCAACTCCGCGCACTGACAGCAAGAGCAAGCCGCGC 2280
Oy 3234 cgcgtgctatgctcgaacacgaacccagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3293
Db 2281 CGCTGGCTATGATCGGAACCGAGCCAGCGCGCGCTACTCAACTCGTGGCAGCTGATC 2340
Oy 3294 taagaagaagcgtacgaactgatacaccagcaatgaagaagccgcgcgcgcgcgcgcgcgc 3353
Db 2341 TACGAGAGAGGGCTACGACATGATCAGCAATGAAGCCCGCACCCCAATACCTCGAC 2400
Oy 3354 cagcttaacaaagaagaagaactgtagagcgaacgctagaagcgaagcgaacgaactagcga 3413
Db 2401 CAGCTTACCAAGAAAGAACTGATGAGCGACGACTAGACAGCGCAACCTACGCA 2460
Oy 3414 cccgagatcttcgaagcgtatgacgcgaagcgaacccgcgcgcgcgcgcgcgcgcgcgcgcgc 3473
Db 2461 CCGAGATGATCTTGCAGCTGATGCGGAGACCGGAAACCGGCTGCTGGCCAGCTACGTTGC 2520
Oy 3474 atggagaagatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3533
Db 2521 ATGGGAAGATGAAGCGCGCTGATCCCGCGAGGCTGCGCACTGGCAGGCGAGGATTCGC 2580
Oy 3534 caccgcgacgaaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3555
Db 2581 CACCCGGACGAGGTGCGCGCC 2602

```

RESULT 4

BLPBLAB 3159 bp DNA 1linear BCF 20-MAR-1997
 LOCUS B.Linens plasmid pBL-a8 DNA for gene encoding Rep protein and 2
 DEFINITION ORF's.

```

ACCESSION Y11902
VERSION Y11902.1 GI:1903028
KEYWORDS ORF2; ORF3; plasmid; rep gene; rep protein.
SOURCE Brevibacterium linens.
ORGANISM Brevibacterium linens.
REFERENCE 1 (bases 1 to 3159)
AUTHORS Leret,V.
JOURNAL Thesis (1995) Universite de Caen I
REFERENCE 2 (bases 1 to 3159)
AUTHORS Leret,V.
JOURNAL Direct Submission
TITLE Submitted (14-MAR-1997) V. Leret, Universite Rennes I, CNRS
JOURNAL UPRE-A6026, Campus de BEAULIEU, Bat 14. Av. du General Leclerc,
35042 Rennes Cedex, FRANCE
FEATURES
     source
         location/Qualifiers
             1..3159
                 /organism="Brevibacterium linens"
                 /plasmid="pBL-a8"
                 /strain="ATCC 19391"
                 /db_xref="taxon:1703"
                 complement(1..516)
                 /gene="ORF3"
                 complement(<1..516)
                 /gene="ORF3"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="CAA72652.1"
                 /db_xref="GI:1903031"
                 /db_xref="SPRMBL:p97025"
                 /translation="MTDDPSGPTRAEAAARDLADNGAVTARAREAAVRMAVASAT
                 AKAMDAASEDDHRTPAYDEVTGALTAIWADAYAAALAAATPEDDQALRSVDALRV
                 EVELADVEIVAEERDRTATQADTLKEIAKAEKWEKSTEAKEHAALSAVRDH
                 DKLINERTITLE"
                 17..283
                 /gene="ORF2"
                 17..283
                 /gene="ORF2"
                 /function="Involved in plasmid copy number regulation"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="CAA72652.1"
                 /db_xref="GI:1903029"
                 /db_xref="SPRMBL:p97023"
                 /translation="MRSLSISCSWRSLASASCLASVDHFSASALASASFLRVSAM
                 VAVSRSASTVSTSAVSPSTSRASSTSRNMSRSGVTAARAAR"
                 1298..2251
                 /gene="REP"
                 1298..2251
                 /gene="REP"
                 /function="necessary for plasmid pBL-a8 replication"
                 /codon_start=1
                 /transl_table=11
                 /product="putative Rep protein"
                 /protein_id="CAA72653.1"
                 /db_xref="GI:1903030"
                 /db_xref="SPRMBL:p97024"
                 /translation="MSTASTETWGMMLPLCPWPLTICSRAFRTSRHNALDVRTIEA
                 NPQSLNLVVDIDHDPAHAACGTEKPOPNVAVENPANGAHAVMALAEPTRTREY
                 ARRKPLAVAAVTEGLRBSVIDGKGYSGILTKNPRHEOWEASMLPHILYNLDELTPHL
                 TVSDPMPDSMORTKRNKPYGLGRNCTLEETVAVMDVRYRCVPSGSAATNPREDRDLD
                 KORSISCARVTEFSALEPASEIRATIRSYKWTITRYTGWLDSTRTSEKSAAYHRN
                 TGRGGGLKGIISAKLAARPPASALMROSAMATRERIPVMG"
BASE COUNT 633 a 1037 c 929 g 560 t
ORIGIN

```

Query Match 3.48; Score 119.6; DB 1; Length 3159;
 Best Local Similarity 48.2%; Pred. No. 5; 3e-11;
 Matches 552; Conservative 0; Mismatches 579; Indels 15; Gaps 7;


```
CDS
/gene="repa"
2696..3604
/gene="repa"
/codon_start=1
/transl_table=11
/product="Replication protein"
/protein_id="BAB17919.1"
/db_xref="GI:11182360"
/translation="MATTTVDQEQEVMVLRPPLASDDILAAGIYRMARPALGVRYIE
VNPQALSNLIVDCDHPDAMRAVMDRHMLPVAIYENPDNGHVAHVMLEAIPTE
VAHRKPIAAVAATBELRRSYGSDASAGLITRNPENHPAMNTWCTDMLYRLAELTH
LDAAGLMPAPSWRRRRRRRPPVGLGRNCAIFETARTAYADARIORRHEYPRAEDSD
LHAVALSTVEALNAGTSEPLPAREANGIASIHRWITHTFTGWDISHVTNEATFTIO
STRGRGAGKARRARRASITDWEA"
3604..3948
/gene="repB"
3604..3948
/gene="repB"
/codon_start=1
/transl_table=11
/product="Replication protein"
/protein_id="BAB17920.1"
/db_xref="GI:11182361"
/translation="MADVOHRYKRRCTARAEARVCASTRQARWTSIREETITOKA
VEREIRAKIKDGHGTWGETSRHFGIAKTTAQRARRARRERAEKAEAEALR
PILFEQEQGSA"
6132..6677
/codon_start=1
/transl_table=11
/product="Resolvasase"
/protein_id="BAB17921.1"
/db_xref="GI:11182362"
/translation="MSITGVARYSTAEOHQAQIDALKAAGADPIYIDKASGATMARP
OMRACIAGSGPGLMVTIRIDLRGSLADLVATDLAGRGVGFSLAEGLDTSPGG
LALFOMAGVFAEERROLIARTRPDGLAARARRHRRGRPALTAEQOASARQMRAGA
SLGSTRILIGVSTRITRTVT"

BASE COUNT      1321 a      2495 c      1993 g      1059 t
ORIGIN

Query Match      3.3%; Score 118.2; DB 1; Length 6688;
Best Local Similarity 53.2%; Pred. No. 8.1e-11;
Matches 327; Conservative 0; Mismatches 273; Indels 15; Gaps 3;

QY 280 cgttcgagagcgtgtccctcgagagcgtgtgtccgacgacgctgtgctcagcagaga 339
    |||||
Db 2709 CGGTCCATGAGCAAGTGGAGACAGGTGTGGCTGCCGCCCTGTGGCTCCGACACAGCA 2768

QY 340 agtctggggcgtacgagcgtgtgtgagagggcgtgtgagctgtcctacacgaag 399
    |||||
Db 2769 TGGCAGCGGGCATCTACCGGATGGCCCGCTCGGGCTGGGGGCTCGATACATCGAGG 2828

QY 400 cgaaccggtgtgtatgcattcctgtgtacacgagtcgagatgtcgtgagtgact 459
    |||||
Db 2829 TCAACCCCAAGCCATAGCAACCTCTCGTGTGCTGACTCGACACCCGACGCTGCCA 2888

QY 460 gggcgcgagacgtcgtgtgtgtcctcaccgtctctacg---tgtcatgaacggtgtca 516
    |||||
Db 2889 TGGCGCGCGGTGTGGAGACCGCAGACGTGGCTGCCCAAGCCATGTGAGAGAACCCGACA 2948

QY 517 cgaaccagcagacatcgtctatgtccttgaagaacctgtgtgtctcagcagatgcgcgc 576
    |||||
Db 2949 ACGGCCACGCCACGCGTGTGGCCCTGGAGACGACCATCCCGACGACGAGTACGCC 3008

QY 577 ggcgaagcgtatcaactcgtcgtccgcgtgcgagcgggctcattgtgaagttctgcgcg 636
    |||||
Db 3009 ACCGCAAGCCCATCGCTACGCGCGCCGCTACCGAGGGGCTTCGCGATCCGTCGACG 3068

QY 637 ggcgtatcctatcagcagcagatcacaagaacccgctcgaagcagcccatgtgcagcc 696
    |||||
Db 3069 GAAGACGCTCTCTACGCGCGGCTGATTCACAAAGAACCCGGAACACCCGCTGGAGACCA 3128

QY 697 tctgtgggccccgcgaagcgc-----gctctcagagctgtgcgcctcgcacacacccctcg 750
```

```
Db 3129 CCTGGTGCACCGACACACCTCTACCGCGCTGGCGGAGCTCGACACCCACCTGTGATCCCGCG 3188
    |||||
QY 751 agcagatccagcagctcgcgagagcagggaacccgcttcgaacgtaccgcgatcaacgcg 810
    |||||
Db 3189 GCCATGACCCCGCCCGCTCTCGGCGACGACCCCGCCGCGACCCGCTGCGCC----- 3242

QY 811 tcggcgcgaacgtacacctgttcgacacacccgcgcatgtggcatalcagcggcctccgcg 870
    |||||
Db 3243 TGGCGCCCACTGGCGCATCTTCGAGACCGCGCGCACCTGGCGCTTACCGGACGCGCGCC 3302

QY 871 actcctggggcgccgc 885
    |||||
Db 3303 GCATCCGACCAACAGCC 3317

RESULT 6
PUE250233
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MIDDLE
REFERENCE
AUTHORS
TITLE
FEATURES
source

PUE250233
6868 bp DNA circular BCT 05-MAY-2001
Propionibacterium jensenii plasmid pLME106 ORF1, ORF2, ORF3, ORF5,
ORF6 and ORF7, ORF8, ORF9 and pna.
AJ250233
AJ250233.1 GI:7635898
invertase; ORF1; ORF10; ORF2; ORF3; ORF5; ORF6; ORF7; ORF8; ORF9;
Pna gene; theta replicase.
Propionibacterium jensenii.
Propionibacterium jensenii.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;
Propionibacterium.
1 (bases 1 to 6868)
Miescher,S., Stierli,M.P., Teuber,M. and Melle,L.
Propionidin SMI, a bacteriocin from Propionibacterium jensenii DFL:
isolation and characterization of the protein and its gene
Syst. Appl. Microbiol. 23 (2), 174-184 (2000)
20383893
2 (bases 1 to 6868)
Stierli,M.P., Melle,L. and Teuber,M.
Molecular analysis of plasmid pLME106, a plasmid isolated from
Propionibacterium jensenii
unpublished
3 (bases 1 to 6868)
Stierli,M.P.
Direct Submission
Submitted (12-OCT-1999) Stierli,M.P., Laboratory of Food
Microbiology, LFO G16, ETH Zurich, Institute of Food Science,
Schmelzbergstrasse 9, CH-8092 Zurich, SWITZERLAND
Location/Qualifiers
1..6868
/organism="Propionibacterium jensenii"
/plasmid="pLME106"
/strain="DF1"
/isolate="Raw Milk"
/db_xref="taxon:1749"
/country="Switzerland"
155..388
/note="ORF8"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB88394.1"
/db_xref="GI:7635899"
/db_xref="SPTREML:091459"
/translation="MRTAVSVKSEEGRLDELARATGSKSYLREALEHDIDQVY
HDVAIRLADVAVAGRAATYSADEVDDIIGLDD"
372..635
/note="ORF10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC38044.1"
/db_xref="GI:1392351"
/translation="MAWTIEYTPAVKALRKLDRQAQRITAYIRIETIGLIDPHQRKQ
```


Db	4871	GACAAAGGCGCTCATATGTTGGCCGCTAAGGCTGGCATCCGACCGGAATTCGCTGCCGGG	4930
Qy	1147	cccata-----cccgtaaacagttcttgaggggtctctcatgac	1185
Db	4931	GTGCAAAATTTCAGAGAGATGACGCTCGGATTCAGAGGAGCGAGACTCGATGCACATGAC	4990
Qy	1186	cacacggaaggtctcccccgaacggtctacgcatcgccgtctgcgcgaaaaagctcg	1245
Db	4991	GCGTGAAATGAGAACCTTCGCAATGGGATGAGTGGCGGACCTTCCGCGGAAGGTCCG	5050
Qy	1246	tgctccgagtcacacgltcaagcgttgatctccgacgacgacgagatctggtcccg	1305
Db	5051	GGTAGGAGAGAAAGCATTCGCTTGACCTCGAGACCGCGGAGATTATCTAAGCCG	5110
Qy	1306	cgttgcgcgacgacgacgagatctgagctcgctcggtcggaaggtcag-aggatcgctg	1364
Db	5111	TGCCCCAACAGCGGCGGTGGCGGATTCGTGAGCTGCGGAAACGGCGCTGCAATGCTG	5170
Qy	1365	cgattgctgcgcgaggttcggttcggtgacacctgacgtacgctgacgaagaatc	1424
Db	5171	CGATGCGCGGTGAGTGTGAGTCTCGTGGCACCGTCCACAAACGGCTCAAGACACAC	5230
Qy	1425	gaactgacg 1433	
Db	5231	CCGAGGAGCG 5239	
RESULT	8		
LOCUS	AF312210	5403 bp	DNA
DEFINITION	Rhodococcus erythropolis plasmid pN30 replication protein and putative DNA-binding protein genes, complete cds.		
ACCESSION	AF312210		
VERSION	AF312210.1	GI:11095364	
KEYWORDS			
SOURCE	Rhodococcus erythropolis.		
ORGANISM	Rhodococcus erythropolis.		
REFERENCE	Bacteria, Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.		
AUTHORS	1 (bases 1 to 5403) Ryabchenko, L.E., Novikov, A.D., Golyshin, P.N. and Yanenko, A.S.		
TITLE	Rhodococcus erythropolis plasmid pN30 putative replicase (ORF1), and putative DNA-binding replication protein (ORF2)		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 5403) Ryabchenko, L.E., Novikov, A.D., Golyshin, P.N. and Yanenko, A.S.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-OCT-2000) Biodegradation Genetics Laboratory, Institute of Genetics and Selection of Industrial Microorganisms, 1st Dorozhny Pr., 1, Moscow 119545, Russia		
FEATURES	Location/Qualifiers		
source	1..5403		
	/organism="Rhodococcus erythropolis"		
	/plasmid="pN30"		
	/db_xref="taxon:1833"		
CDS	1160..2047		
	/note="ORF1, putative; similar to putative Rhodococcus rhodochrous PR4 putative replication protein (repA), Rhodococcus rhodochrous PKA2 putative theta replicase (ORF2), Rhodococcus erythropolis pRAJ2600 putative replication protein (repA), Mycobacterium fortuitum pAJ5000 putative replication protein (37K)"		
	/codon_start=1		
	/transl_table=11		
	/product="replication protein"		
	/protein_id="AAG29854.1"		
	/db_xref="GI:11095365"		
	/translation="MKSATGVLSDROWADAVLPHREPATNNTLRGQSRSDALA MRVESHPLALGSIYVDCDHDAAMRAFEQSDPAPNVAQSPSGAHTGWLGP NVHRTSARLTPLRYAHRIETGLKISVGDFAYGGOLKNPIDPMWETVBPATPYL CHVATLHTRPROMRPDRAVAGTGRVNTMDARRRAAYRQMMOHNRGTRDMDHVLOH RAATTEPTTPTPEFTEVRAATQASISKMTWRNTEEQYARQAHQKQKGAATTLAKOE AVNNARKTDEHTRERAIL"		

Db	2195	CGTGTGACCTGGGGAAGCAGGGGTTGAAGTACCGGGGAATCGCCGAGGAGTGAACCTCT	2254
Oy	1389	ccgtgtgacccgtgtgacacgcgtcgacacgaatcgactgacg	1433
Db	2255	CGACCGGAGATCGTCGGCGATTACTGCAACGACGCCGCAAGGACG	2299
RESULT	9		
AB040101			
DEFINITION	AB040101	2582 bp	DNA
LOCUS	Rhodococcus rhodochrous plasmid pRC4 repA, repB genes for		linear BCT 06-APR-2001
ACCESSION	AB040101		
KEYWORDS	AB040101.1	GI:7262572	
SOURCE	DNA-binding replication protein; replication protein.		
ORGANISM	Rhodococcus rhodochrous (strain:IFO3338) plasmid:pRC4 DNA.		
REFERENCE	Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:		
AUTHORS	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.		
TITLE	1 (sites)		
JOURNAL	Hirasawa, K., Ishii, Y., Kobayashi, M., Koizumi, K. and Maruhashi, K.		
REFERENCE	2 (bases 1 to 2582)		
AUTHORS	Ishii, Y., Hirasawa, K. and Maruhashi, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-MAR-2000) Kenji Maruhashi, Petroleum Energy Center,		
FEATURES	Bio-Retining Process Laboratory, Sodaeshi-cho 1900, Shimizu,		
source	Shizuoka 424-0037, Japan (E-mail:k.maruhashi@pecj.or.jp,		
	Tel:+81-543-67-9550, Fax:+81-543-67-9552)		
	location/Qualifiers		
	1. 2582		
	/organism="Rhodococcus rhodochrous"		
	/plasmid="pRC4"		
	/strain="IFO3338"		
	/db_xref="taxon:1829"		
	1142. 2062		
	/gene="repA"		
	1142. 2062		
	/gene="repA"		
	/note="putative		
	similar to putative theta replicase from Rhodococcus		
	rhodochrous PKA22 and RepA proteins from Mycobacterium		
	fortuitum PAL5000 and related plasmids."		
	/codon_start=1		
	/transl_table=11		
	/product="replication protein"		
	/protein_id="BA092818.1"		
	/db_xref="GI:7262573"		
	/translation="MDMSGSLGDMWOLMLPLPMTADDLLGVYMPRODALDRRY		
	LEAPDQSLNLVYDVDPDAALALASAGNHPNPAIVENPRGHAHAYALTEPTT		
	RTETARRKPPPLAAVNBEGLEKRAVDGDAASGLMTKNPTSHAMPTWTHATRLADI		
	EHDIQKHPPRMROSKRRREDPVGLGNCMLFTARTWAVRELRCHWGEDEGKAI		
	QVEADLIANAASEPLPVSSEVRAIAASIHRTVTSKRMWADGPAYTEFAVIOASRGR		
	KMTCKREANRRATRTKXYDRDLVRKEATDGS"		
	2052. 2333		
	/gene="repB"		
	2052. 2333		
	/gene="repB"		
	/note="repB"		
	/note="putative		
	similar to putative DNA-binding replication protein from		
	Rhodococcus erythropolis pRAJ2600"		
	/codon_start=1		
	/transl_table=11		
	/product="DNA-binding replication protein"		
	/protein_id="BA092819.1"		
	/db_xref="GI:7262574"		
	/translation="MGAETPARPRTAREVNERIGASPRVRIIAPRASYEARAAE		
	RRKQVLEIRASGMKRLRETAIVAGSVGGTILHARTKTESKAEGAMA"		

[illegible]


```

CDS
    complement(1503..2297)
    /codon_start=1
    /transl_table=11
    /product="neomycin-kanamycin phosphotransferase"
    /protein_id="AAC53671.1"
    /db_xref="GI:993049"
    /translation="MIEDGGIHAGSPAAWERLFGYDWAQQTIGCSDAAVRLSAQGH
    PVLEKTDLSGALNEQDEARLISMLATTGPGCAAVLDVYDEARQDWILLGEVGGODI
    LSHLPAEKVSIINADMRRLHTLDPATCGPDQAKHIEARLRMEGLVGGDDLEED
    OHQGLAPAEILFARLKRAMPDEGDELVYTHGDCALNINWENRFGFIDCGRLGVADRRL
    QDIALATRIADIEEGGEMADRFVLVYHIAADPSQRIAFVRLDDEFF"
    2665..5252
    /organism="Plasmid PAL5000"
    /note="EcoRV/hpaI from PAL5000; minimal mycobacterial
    replicon from PAL5000, Genbank Accession Number M23557"
    /db_xref="taxon:2522"

BASE COUNT
    1054 a 1596 c 1518 g 1084 t

ORIGIN

```

Query Match	2.3%	Score 82.2	DB 12	Length 5252
Best Local Similarity	47.8%	Pred. No. 0.00013		
Matches 274	Conservative 0	Mismatches 293	Indels 6	Gaps 1

[illegible]

RESULT	12
I31847/c	
LOCUS	I31847
DEFINITION	Sequence from patent US 5583038.
ACCESSION	I31847
VERSION	I31847.1
KEYWORDS	GT:1822638
SOURCE	Unknown.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	SOURCE	BASE COUNT	ORIGIN
Unknown.	1 (bases 1 to 6407)	Stover, C.K.	Bacterial expression vectors containing DNA encoding secretion signals of lipoproteins	Patent: US 5583038-A 1 10-DEC-1996;	Location/Qualifiers	1..6407	1399 a 1800 c 1770 g 1438 t	
						/organism="unknown"		

Query Match	2.3%	Score	82.2	DB	6	Length	6407
Best Local Similarity	47.8%	Pred	No	0.00013			
Matches	274	Conservative	0	Mismatches	293	Indels	6
						Gaps	1

[illegible]

LOCUS	131848	6407 bp	DNA	linear	PAT 06-FEB-1997
DEFINITION	Sequence 2 from patent US 5583038.				
ACCESSION	131848				
VERSION	131848.1	GI:1822639			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 6407)				
AUTHORS	Stover,C.K.				
TITLE	Bacterial expression vectors containing DNA encoding secretion signals of lipoproteins				

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2002, 11:26:07 ; Search time 532.75 Seconds
(without alignments)
11456.839 Million cell updates/sec

Title: US-09-720-583A-1

Perfect score: 3555
Sequence: 1 gtcgaccttaccagccgccc.....ccgcacgaggtgcgccc 3555

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3555	100.0	3555	21	AAZ49691
2	82.2	2.3	1382	19	AAV69313
3	82.2	2.3	1463	18	AAV69312
4	82.2	2.3	6407	19	AAV64413
5	80.6	2.3	4120	13	AAQ21502
6	80.6	2.3	4120	18	AAV64421
7	80.6	2.3	4296	17	AAV28272
8	80.6	2.3	4352	17	AAV28273
9	80.6	2.3	4505	17	AAV28276

10	80.6	2.3	6047	17	AAV28278	Plasmid pMH30-lux
11	80.6	2.3	6171	17	AAV28279	Plasmid pMV261-lux
12	77.4	2.2	4120	17	AAV28281	Plasmid pMV206 seq
13	68.6	1.9	4119	14	AAQ41331	pMV206 - cassette
14	68.6	1.9	6407	14	AAQ41316	pMV101 - a deletion
15	58.6	1.6	4811	13	AAQ31743	pMV101 - a deletion
16	58.6	1.6	4811	13	AAQ31727	pMV101 - a deletion
17	56.2	1.6	4674	21	AAV14666	Nucleotide sequenc
18	56.2	1.6	4725	21	AAV14665	Nucleotide sequenc
19	56.2	1.6	4737	21	AAV14668	Nucleotide sequenc
20	56.2	1.6	4767	21	AAV14667	Nucleotide sequenc
21	56.2	1.6	4818	21	AAV14669	Nucleotide sequenc
22	52.4	1.5	65140	22	AAV17184	Streptomyces nous
23	52.4	1.5	125401	22	AAV17186	Streptomyces nous
24	50.2	1.4	1524	20	AAV209259	Rice RPC213 CDNA
25	50.2	1.4	1524	22	AAV74817	Rice CDNA sequence
26	50.2	1.4	4466	21	AAV14663	Nucleotide sequenc
27	50.2	1.4	4478	21	AAV14661	Nucleotide sequenc
28	50.2	1.4	4547	21	AAV14664	Nucleotide sequenc
29	50.2	1.4	4571	21	AAV14662	Nucleotide sequenc
30	50.2	1.4	77536	21	AAV14651	Nucleotide sequenc
31	49.8	1.4	1881	24	AAV18436	Contig 109 DNA enc
32	49.4	1.4	1350	21	AAV58844	Mitomycin gene clu
33	49.4	1.4	18034	21	AAV58841	Complete Mitomycin
34	48.2	1.4	833	15	AAV64203	snab gene encoding
35	48.2	1.4	1896	21	AAV64297	S. coelicolor YesM
36	48.2	1.4	5392	15	AAV64201	Sequence compisn
37	47.4	1.3	27541	22	AAV17185	Streptomyces nous
38	47.4	1.3	58857	22	AAV58871	Nucleotide sequenc
39	47	1.3	5824	22	AAV54491	CDNA encoding nove
40	47	1.3	5824	22	AAV51846	Human polynucleoti
41	47	1.3	5824	22	AAV51846	Human polynucleoti
42	47	1.3	77536	21	AAV14651	Nucleotide sequenc
43	46.6	1.3	11220	21	AAV287298	S. venezuelae macr
44	46.6	1.3	36778	21	AAV287318	S. venezuelae plk
45	46.6	1.3	37948	21	AAV287285	S. venezuelae plk

ALIGNMENTS

RESULT 1	AAZ49691	standard; DNA; 3555 BP.
ID	AAZ49691	
XX	AAZ49691	
AC	AAZ49691	
XX	07-APR-2000	(first entry)
DT	07-APR-2000	(first entry)
XX	Proprionbacterium	plasmid LMG 16545 DNA.
DE	Proprionbacterium	plasmid LMG 16545 DNA.
XX	Proprionbacterium	LMG 16545; vector; plasmid; antigen; vaccine; enzyme;
KW	Proprionbacterium	LMG 16545; vector; plasmid; antigen; vaccine; enzyme;
KW	nutritional factor; growth factor; clotting factor; antimicrobial; drug;	
KW	hormone; vitamin B12; animal feed; lactic acid bacteria; foodstuff;	
KW	cheese; cyclic; circular; da.	
XX	Proprionbacterium	freudenreichii LMG 16545.
OS	Proprionbacterium	freudenreichii LMG 16545.
XX	Proprionbacterium	freudenreichii LMG 16545.
FH	key	Location/Qualifiers
FT	CDS	273..1184
FT	FT	/tag- a
FT	FT	/product- "Proprionbacterium LMG 16545 protein-1"
FT	FT	/note- "Encoded by ORF-1"
FT	FT	1181..1438
FT	FT	/tag- b
FT	FT	/product- "Proprionbacterium LMG 16545 protein-2"
FT	FT	/note- "Encoded by ORF-2"
PN	MO9967356-A2.	
XX	29-DEC-1999.	
PD	29-DEC-1999.	
XX	25-JUN-1999.	99MO-EP04416.
PF	25-JUN-1999.	99MO-EP04416.

QY	535	tctatgcttgaagaacacccgtgtgtctatgaacgatgcgcgcgcgcgcgcgcctatcaac	594
Db	2612	TGTTGGGCACTCAACGCCCTCTGTTCCACGCACCAATATACGGCGGCTTAAGCCGCTGCAT	2553
QY	595	tgtctgcgccgcgtctgcagcaagcgctatgtgcagcttctcgcgcgcgatgatcctactcagcgc	654
Db	2552	ACATATGCGGCGGTGGCGAAGGCGCTTTCGGGCGCGCTGCACGCGACCGCACTTACTG	2493
QY	655	acgggatctcaagaagaacccgctcaagaacgcgcgcgcgcgcgcctcttggggcccgacagc	714
Db	2492	GCCCTCATGACCCMAAACCCCGCGGCACATCGCTGGGGAAACGAAATGGCTCCACTCAGATC	2433
QY	715	cgctctcagagctgtgcgcgcctcgcacaacccctgcagagatccaagcactgcgcgcgcgc	774
Db	2432	TCTTACACACTCAGCCACATGATGAGGCCGAGCTCGCGCGCAACATGCACCGCGCGTGGC	2373
QY	775	caagggaaccgcgctcgcagacgtlaccgcgatlcaacggtctgcgcgcgaacgltcacccgttgc	834
Db	2372	GTCAGACAGACACGCTACAAAGGCGGCTCCGACGCGCGCTAAGGGGGAAATTGCGACTGTTCG	2313
QY	835	acaccaccgcgatgtggcgtaccatccgcgcgtcc	867
Db	2312	ATTCCGTGAGGTTGTGGGCTATCTATCTCGCGCC	2280
RESULT	5		
AAQ21502			
ID	AAQ21502	standard; DNA; 4120 BP.	
XX	AAQ21502;		
AC			
XX	03-JUN-1992	(first entry)	
DT			
XX			
DE		Vector pMW206 for cloning genes in E.coli and mycobacteria.	
XX			
KW		Polymerase chain reaction: mycobacterial promoter; kanamycin;	
KW		kanR; aph gene; BCG; Bacille Calmette-Guerin;	
KW		site-specific integration; ss.	
XX			
OS		Escherichia coli.	
OS		Mycobacterium smegmatis.	
XX			
XX			
FT	Key	Location/Qualifiers	
FT	CDS	121..936	
FT		/*tag= a	
FT		/product= kanR	
FT		/note= "Kanamycin resistance gene"	
FT		1..1219	
FT		/*tag= b	
FT		/note= "kanamycin resistance cassette"	
FT		1220..2011	
FT		/*tag= c	
FT		/note= "E.rep cassette"	
FT		2012..2025	
FT		/*tag= d	
FT		/note= "MluI-NotI linker"	
FT		2026..3924	
FT		/*tag= e	
FT		/note= "M.rep cassette"	
FT		3925..4016	
FT		/*tag= f	
FT		/function= Multiple cloning site	
FT		/note= "contains 16 unique restriction sites,	
FT		stop codons in all 3 reading frames and	
FT		a transcription terminator"	
XX			
XX			
PN	WO9201783-A.		
PD	06-FEB-1992.		
XX			
XX	09-JUL-1991.	91WO-US04833.	
XX			

PR 16-JUL-1990; 90US-0553907.
XX
XX (YESH) EINSTEIN A COLLEGE.
PA (UYPL-) UNIV OF PITTSBURGH.
XX
XX
PI Jacobs WR, Hatfull G;
XX
DR WPI, 1992-064943/08.
DR P-PSDB; AAR20993.
XX
XX DNA site-specific integration into Mycobacteria - useful as
PT adjuvant in vaccines and as therapeutic agent for malaria,
PT influenza, herpes and human immunodeficiency virus
XX
PS
XX Example 3; Fig 31; 82p; English.
XX
CC Casettes of all the components necessary for plasmid replication
CC in E.coli and mycobacteria (E.rep and M.rep, respectively) and for
CC selection of transformants (KanR) were constructed using the PCR
CC technique. A multiple cloning site was synthesised. The cassettes
CC were constructed to allow directional cloning and assembly into a
CC plasmid where all transcription is unidirectional. Plasmid pmv206
CC was derived from plasmid pmv204 which was made up of all the
CC cassettes. It was modified by the insertion of an MluI-NotI linker
CC between the M.rep and E.rep cassettes to facilitate removal of
CC M.rep in further constructions.
CC See AAO21568-021573 for the PCR primers used in the construction of
CC the individual cassettes.
XX
XX Sequence 4120 BP; 923 A; 1127 C; 1163 G; 907 T; 0 other;

Query Match	Similarity	2.3%	Score 80.6	DB 13	Length 4120
Best Local	Similarity	47.6%	Pred. NO. 1.3e-09		
Matches	273	Conservative	0	Mismatches	294
				Indels	6
				Gaps	1
QY	301	agaagctgtgtgcacacgaacagccgtctgtgtcagccgacgaagatcttgggtgcgtacacgaacg	360		
Db	2382	agctctgtgtctgcgtactgtgcgcgtctgtgcaaggaagactgtgtctgaagggtactctacgcgc	2441		
QY	361	tgaatctgtgcagaggtgcgtctgtgaagcttgccttacaatctgaagcgaaccgtttgtgtcagt	420		
Db	2442	aaagcgcgcgtctgcgcgccttaagccgcgcgtctacatctgaagcgaaccacaacagcgtctgcaa	2501		
QY	421	ccttgtatcatcacctgaagatgagatgtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	480		
Db	2502	acctgtctgtctgt	2561		
QY	481	tgtcttcaacgcgtctctacagtgcttccatgt-----aacccgtgtctacagacacacacacag	534		
Db	2562	gtctccatccgtctgtgcacaaagcgtatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2621		
QY	535	tctatgtcctttgaagaaccctgt	594		
Db	2622	tgtgtgtgcactcaaacgccctcttccacagcacgtaatacgtgcgcgtgtgtgtgtgtgtgtgtgtgt	2681		
QY	595	tgtctgtccgcgtctgtgaagaggtgcctatgtcgaagcttctctgtgtgtgtgtgtgtgtgtgtgtgt	654		
Db	2682	acatgtgtgcgtgtgtgcgaagaagcctctctgtgtgcgcgcgtctgtgtgtgtgtgtgtgtgtgtgt	2741		
QY	655	aacggtatcaacaagaagaccgcgtctcaacacacgcgcctcaatgtgcacccctctctgtgtgtgtgtgtgt	714		
Db	2742	gtctcatgtacaaaaccgccgcgcacacatctgcctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2801		
QY	715	cgctctcaagagctgtgcgcgcctctgtgcacaaaccctctgtacgtgaagctccacgcatgtgtgcgaag	774		
Db	2802	tctacaacactcagcacacacatcgaagtcgcgaagctctgtgcgaatactgtccacgcgcgcgtgtgc	2861		
QY	775	caggtgaacccgcgtctgtgcgaagctcaacacgtatcaagcgttctgtgtgtgtgtgtgtgtgtgtgtgt	834		
Db	2862	gtcagtcagaaacacgtatacaaaaacgtgtctgcgaacgcgcctgtgtgtgtgtgtgtgtgtgtgtgt	2921		
QY	835	acacacaccgcgatgt	867		

```

Db      2922 atccgcacagtgctggtgagccatcgtccgcc 2954
|      | | | | | | | | | | | | | | | | | |
RESULT 6
ID      AAT64421
AC      AAT64421;
DT      28-MAY-1997 (first entry)
DE      Plasmid pMV206.
KW      Plasmid pMV101; Mycobacteria; lipoprotein secretion signal sequence;
        antigen; bacterial vaccine; Lyme disease; Borrelia burgdorferi;
        surface protein; primer; amplify; PCR; polymerase chain reaction; ss.
OS      Synthetic.
XX
FH      Key
FH      misc_feature
FT      1..1219
        /tag= a
        /note= "NheI/SpeI kan cassette"
FT      121..936
        /tag= b
        /note= "kan gene"
FT      1220..2011
        /tag= c
        /note= "E.rep cassette"
FT      2012..3920
        /tag= d
        /note= "M.rep cassette"
FT      3925..4120
        /tag= e
        /note= "Synthetic multiple cloning site"
XX
XX      US5583038-A.
XX
XX      10-DEC-1996.
XX
XX      21-OCT-1991; 91US-0780261.
XX
XX      17-NOV-1992; 92US-0977630.
XX      21-OCT-1991; 91US-0780261.
XX      (MEDT-) MEDIMMONE INC.
XX
XX      Stover CK;
XX      PI
XX      WPI; 1997-042315/04.
XX
XX      Mycobacteria expressing secretion signal of lipoprotein and
XX      heterologous antigen, esp. outer surface protein A or B of Borrelia
XX      burgdorferi - are used in the form of a live bacterial vaccines
XX      against Lyme disease
XX
XX      Example 1; Fig 12; 122pp; English.
XX
XX      This sequence represents the plasmid pMV206. This plasmid was
XX      used in the construction of a series of vectors used to produce
XX      the recombinant Mycobacteria of the invention. The recombinant
XX      mycobacteria of the invention are transformed with DNA encoding
XX      a polypeptide which comprises a lipoprotein secretion signal
XX      sequence and an antigen (Ag) heterologous to the mycobacteria.
XX      The lipoprotein secretion signal causes the Ag to be produced as
XX      a lipoprotein. The mycobacteria may be used in the form of a live
XX      bacterial vaccines against Lyme disease, where the bacteria express
XX      a surface protein of Borrelia burgdorferi, the causative agent of
XX      Lyme disease.
XX
XX      Sequence 4120 BP; 922 A; 1125 C; 1166 G; 907 T; 0 other;

```

Query Match	Similarity	2.3%	Score 80.6	DB 18	Length 4120
Best Local	Similarity	47.6%	Pred. No. 1.3e-09		
Matches 273	Conservative	0	Mismatches 294	Indels 6	Gaps 1
QY	301	agagctgtgcgcgaacgcgcctgtgcgtctacgcggaagactgtgggtatccgcgcacg	360		
DB	2382	agctctgtgcgcgtactgtgcgcctgtgcgcgaagcgagactgtgtctgagggatctaacgc	2441		
QY	361	tgacttcgcgcagagcgctgtgagctgtgccttaccatcgaagaccgcgttgtcatgcagt	420		
DB	2442	aaagccgcgcgtgcgcgtccctagagccgcgcgttaccatcgaagcggaaccacagcgctgtgc	2501		
QY	421	cccttgatcaacgcgtacgtatgtcttgagctgtgcactgcgcgcgcgcgcgcgcgcgcgc	480		
DB	2502	acctgt	2561		
QY	481	tgcttcaacgcgcctcactgtgtctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	534		
DB	2562	gtctccatcgcgt	2621		
QY	533	tcctatgtccttggaagaaccctgt	594		
DB	2622	tggt	2681		
QY	595	tgctgtgcgcgcgcgt	654		
DB	2682	acatgt	2741		
QY	655	accggtatcaaaagaaacccgcctcgaagacccgcctcgtgtgtgtgtgtgtgtgtgtgtgtgt	714		
DB	2742	gctctatgtacccaataaacccgcctcgaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2801		
QY	715	cgcttcaagactgt	774		
DB	2802	tcctaacactcagcaccatcgt	2861		
QY	775	caagggaacccgcgt	834		
DB	2862	gtcagcagagacacgtatcaaaagcgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	2921		
QY	835	acacacaccgcgcgt	867		
DB	2922	atccgt	2954		
RESULT	7				
ID	AAT28272	standard; DNA; 4296 BP.			
XX	AC	AAT28272;			
XX	DT	19-AUG-1996 (first entry)			
XX	DE	Plasmid pMH28 sequence.			
XX	KM	Plasmid pMH28; vector; firefly; luciferase; DNA cassette;			
XX	KM	reporter gene; luminescence; BGL; Mycobacterium			
XX	KM	Mycobacterium tuberculosis; Mycobacterium			
XX	KM	Mycobacterium intracellulare; Mycobacterium			
XX	KM	Mycobacterium scrofulaceum; Mycobacterium			
XX	KM	leprosy; antibiotic; tuberculosis; vaccine; drug			
XX	OS	Synthetic.			
XX	PN	WO9610645-A1.			
XX	PD	11-APR-1996.			
XX	PF	02-OCT-1995; 95WO-US12642.			
XX	PR	03-OCT-1994; 94US-0316950.			
XX					

QY 481 tgccttaacgctctactgctgtccatg-----aacggttcaacgacacacgacacatcg 534
 DB 2585 ggtcccatcgctgcccacacgcatcgctgtgacatcgccgacacgacacacgacg 2644
 QY 535 tctatgcttgaagaacccctgtgtgtctgaacgatcgccgacgacgacgacacac 534
 DB 2645 tctggaactcaacgcccctgttccacgacacgacacgacgacgacgacgacgac 2704
 QY 595 tgcctcccgctcgcgacgacgacgacgacgacgacgacgacgacgacgacgacgac 654
 DB 2705 acatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 2764
 QY 655 accgagatcaaaagaacccgctcgaacgacgacgacgacgacgacgacgacgacgac 714
 DB 2765 gctcctacgacacaaaccccgacacgacgacgacgacgacgacgacgacgacgac 2824
 QY 715 cgcctcagagctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 774
 DB 2825 tctacacactcagcaccatcgaagcgcgacgacgacgacgacgacgacgacgacgac 2884
 QY 775 caaggaaccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 834
 DB 2885 gtcagagacacacgacacgacgacgacgacgacgacgacgacgacgacgacgacgac 2944
 QY 835 acacacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 867
 DB 2945 attcgcacgattgt 2977

RESULT 9

AAT28276
 ID AAT28276 standard; DNA; 4505 BP.

AC AAT28276;

DT 19-AUG-1996 (first entry)

DE Plasmid pMW261 sequence.

XX Plasmid pMW261; shuttle vector; *Escherichia coli*; firefly;
 luciferase; DNA cassette; reporter gene; luminescence; BCG;
 KM *Mycobacterium bovis*; *Mycobacterium tuberculosis*;
 KM *Mycobacterium avium*; *Mycobacterium intracellulare*;
 KM *Mycobacterium kansasii*; *Mycobacterium scrofulaceum*;
 KM *Mycobacterium leprae*; tuberculosis; leprosy; antibiotic;
 KM tuberculostatic; vaccine; drug screening; ss.
 XX Synthetic.

OS W09610645-A1.

PN 11-APR-1996.

PD 02-OCT-1995; 95WO-US12642.

PR 03-OCT-1994; 94US-0316950.

PA (PATH-) PATHOGENESIS CORP.

PI Hickey MJ, Stover K;

DR WPI; 1996-209365/21.

XX Quantification of mycobacteria by luminescence - using bacterial
 PT reporter strains contg. a lux gene, permits the determ. of the
 PT efficacy of anti-mycobacterial and prophylactic compns.

XX Example 1; Page 53-56; 82pp; English.

XX This is the complete sequence of plasmid pMW261. This plasmid is
 CC an extrachromosomal *Mycobacterium-Escherichia coli* shuttle
 CC expression vector. A mutant Pflux firefly luciferase (EC-1.13.12.7)
 CC reporter gene DNA cassette may be inserted downstream from a

CC synthetic hsp70-tac promoter to give plasmid pMW261-lux (AAT28279).
 CC This type of vector may be used in a new method for quantifying
 CC *Mycobacterium* infection in vivo. An animal is injected with e.g.
 CC *Mycobacterium bovis* BCG, etc., transfected with one of these vectors,
 CC and reporter gene expression is detected by luminescence in tissues
 CC without lysis or cell concentration. The method may be used to
 CC determine the efficacy of antibiotic, tuberculostatic and vaccine
 CC compositions.

XX Sequence 4505 BP; 992 A; 1238 C; 1303 G; 972 T; 0 other;

Query Match 2.3%; Score 80.6; DB 17; Length 4505;
 Best Local Similarity 47.6%; Pred. No. 1.4e-09;
 Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

QY 301 agagcttgctgacacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 360
 DB 2382 agctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2441
 QY 361 tgactcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 420
 DB 2442 aaagccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 2501
 QY 421 ccttggtacacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 480
 DB 2502 acctgt 2561
 QY 481 tgccttaacgctctactgctgtccatg-----aacggttcaacgacacacatcg 534
 DB 2562 ggtcccatcgctgcccacacgcatcgctgtgacatcgccgacacgacgacacacgacg 2621
 QY 535 tctatgcttgaagaacccctgtgtgtctgaacgatcgccgacgacgacgacgacgac 594
 DB 2622 tgtggacactcaacgcccctgttccacgacgacgacgacgacgacgacgacgacgac 2681
 QY 595 tgcctcccgctcgcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 654
 DB 2682 acatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 2741
 QY 655 accgagatcaaaagaacccgctcgaacgacgacgacgacgacgacgacgacgacgac 714
 DB 2742 gctcctacgacacaaaccccgacacgacgacgacgacgacgacgacgacgacgac 2801
 QY 715 cgcctcagagctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 774
 DB 2802 tctacacactcagcaccatcgaagcgcgacgacgacgacgacgacgacgacgacgac 2861
 QY 775 caaggaaccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 834
 DB 2862 gtcagagacacacgacacgacgacgacgacgacgacgacgacgacgacgacgacgac 2921
 QY 835 acacacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 867
 DB 2922 attcgcacgattgt 2954

RESULT 10

AAT28278
 ID AAT28278 standard; DNA; 6047 BP.

AC AAT28278;

DT 19-AUG-1996 (first entry)

DE Plasmid pMH30-lux sequence.

XX Plasmid pMH30-lux; vector; firefly; luciferase; DNA cassette;
 KM reporter gene; luminescence; BCG; hsp70-tac promoter;
 KM *Mycobacterium bovis*; *Mycobacterium tuberculosis*; *Mycobacterium avium*;
 KM *Mycobacterium intracellulare*; *Mycobacterium kansasii*;
 KM *Mycobacterium scrofulaceum*; *Mycobacterium leprae*; tuberculosis;
 KM leprosy; antibiotic; tuberculostatic; vaccine; drug screening; ss.

PR 21-OCT-1991: 91US-0780261.
 XX (MEDI-) MEDIMMUNE INC.
 XX
 PI Stover CK;
 XX
 DR WPI: 1993-152187/18.
 XX P-PSDB; AAR34542; AAR34543; AAR34544.
 XX
 PT Expression vector for expressing protein or polypeptide in
 PT Mycobacterium - contg DNA sequences encoding lipoprotein
 PT secretion signal and peptide heterologous to bacteria expressing
 PT fusion protein of lipoprotein heterologous to bacteria
 PS
 PS Example 1; Fig 5; 86pp; English.

This sequence represents that of plasmid pMV101. This is a modified version of plasmid pYUB125, which had 792 bases of the tet gene (inactivated by prior manipulations) deleted. The plasmid represents an expression vector. The HSP61 coding sequence was inserted between the NheI and BamHI sites to form pMV65A. NOTE: - When the proteins encoded by this sequence were decoded, the published sequence was found to contain approx. 65 codons which did not code for the amino acids dictated by the universal genetic code. The sequence also contained many insertions and deletions which upset the open reading frame that was displayed as the encoded proteins. The complementary strand was also found to contain many insertions and deletions that did not have corresponding complementary bases on the coding strand.

Sequence 6407 BP; 1405 A; 1892 C; 1670 G; 1439 T; 1 other;

Query Match 1.9%; Score 68.6; DB 14; Length 6407;
 Best Local Similarity 47.5%; Pred. No. 1.3e-06;
 Matches 272; Conservative 0; Mismatches 294; Indels 7; Gaps 2;

QY 301 agagctggtcgcagcagccgctggtcagccgagaagcttggtggtcagcgcagc 360
 DB 2853 AGCTGTGGCTGCCGTACTGGGCGCTGGCAAGCAGCATCTG-CTCGAGGGGATCTACGCC 2795
 QY 361 tgactcgagagagggcgctggagctgcttacatcgaaagcgaaccgttggtatcgat 420
 DB 2794 AAGCGCGCGCTGGCGCTTAAAGGCGCGCTGACATCGAGGGAACCAAGAGGGCTGGCAA 2735
 QY 421 ccttggtacacccagatgctgagatgctgagatgctgagcgcgcgcgcgcgcgcgcgc 480
 DB 2734 ACCTGCTGCTGGGACGTAAGACCATCCAGACCGCTCCGAGCGCTCAGCGCGGCGG 2675
 QY 481 tgcctcacgctcctcactgctccat-----gaaccgtgtcagacacgcgcgcacatcg 534
 DB 2674 GGTCCCATCCGCTGCCCAACGCGCATGCTGGGCAATGGCGCAACGGCGCAACGACGAC 2615
 QY 535 tctatgcttgaagaacctgtgtgtctgacgagtcgcgcgcgcgcgcgcgcgcgcgcgcgc 584
 DB 2614 TGTGGGGACTCAACGGCGCTGTTCCACGACCAACCAATACGGCGCTTAAGCCCTCGCAT 2555
 QY 595 tgc 654
 DB 2554 ACATGGCGCGCTGCCGGAAGGCTTCGGCGCGCGCTCGACGGCGACCCGATTTACTCAG 2495
 QY 655 accgcgacaaagaacccgctcgaacacccgcatatgcgaacctcttggtggcccccgcgcgcgcgc 714
 DB 2494 GCCTCATGACCAAAAGCGCGCGCATCGCTGGGAAAGGGAATGGTCTCACTCAGATC 2435
 QY 715 cgcctcagagctgctgc 774
 DB 2434 TCTACACACTCAGCCATCCATCGAGCGAGCTGGGGCGGAACGACCCGCCGCTGGCG 2375
 QY 775 cagggagaccgc 834
 DB 2374 GTGACGACGACCAAGTCAAAAGCGGCTCCGACGCCGCTAAGGGCGGAATTTGGCAGCTGTCG 2315

QY 835 acaccaccgc 867
 DB 2314 ATTCGCTCAGGTTGTGGGCTATCGTCGCGGCC 2282

RESULT 15
 AAQ31743/C
 ID AAQ31743 standard; DNA; 4811 BP.
 XX
 AC AAQ31743;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE PMV101 - a deletion mutant of pYUB125.
 XX
 KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
 KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
 KW pertussis; malaria; influenza virus; CTL; herpes virus; ds.
 XX
 OS Mycobacterium.
 XX
 FH Key Location/Qualifiers
 FT misc_difference 2703..2705
 FT /tag= a
 FT /note= "This codon represents 1700 bases missing from the sequence reproduced in the specification"
 FT cds complement (1598..2545)
 FT /tag= D
 FT /product= tet e protein complement (1956..2700)
 FT /tag= C
 FT /product= tet d protein
 FT /note= "precise start posn. unknown as it is contd. in missing portion of sequence."
 FT cds 3661..4474
 FT /tag= d
 FT /product= kan protein
 XX
 PN M09221374-A.
 XX
 PD 10-DEC-1992.
 XX
 PF 01-JUN-1992; 92MO-US05023.
 XX
 PR 06-JUN-1991; 91US-0711084.
 XX
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Dela Cruz V, Stover CK;
 XX
 DR WPI: 1992-433378/52.
 XX
 PT Tetanus vaccination - by provoking an immune response using transformed Mycobacteria
 PS
 PS Example 1; Fig 5; 86pp; English.

This sequence represents that of plasmid pMV101. This is a modified version of plasmid pYUB125, which had 792 bases of the tet gene (inactivated by prior manipulations) deleted. The plasmid represents an expression vector. The HSP61 coding sequence was inserted between the NheI and BamHI sites to form pMV65A. NOTE: - When the proteins encoded by this sequence were decoded, the published sequence was found to contain approx. 65 codons which did not code for the amino acids dictated by the universal genetic code.

Sequence 4811 BP; 1076 A; 1376 C; 1225 G; 1130 T; 4 other;
 Query Match 1.6%; Score 58.6; DB 13; Length 4811;
 Best Local Similarity 47.6%; Pred. No. 0.00032;

	Matches	172;	Conservative	0;	Mismatches	189;	Indels	0;	Gaps	0;
QY	507	aaccggtc	acgacacacgacacacatcgtctatgacctgaagaaacctgtgtgtcgaacc	566						
Db	2642	AATCGCGCAATCGGCGACGACGACGAGTGTGGGACTCAACGCCCTGTTCACGCGACC	2583							
QY	567	gattccgagcgagcgagccatacaacctgtgtcccggtcgaagagggcctatggaac	626							
Db	2582	GAATACGCGCGCGCTAAGCCGCTCGCATACATGGCGGCGTGGCGAAGGCCCTTCGGCGC	2523							
QY	627	gtctcggcgagcgtatgcaatccacgagacggaatcaaaagaaaccgctcagcaccgac	686							
Db	2522	GGCGTCAAGCGCGACCGAGTTACTCAGGCGCTCATGACCAAAAAGCGGCGACATCGCC	2463							
QY	687	catcggaacctctgtgggcccccgacgagcgctctacgaagctgcgccccctgcacacac	746							
Db	2462	TGGGAACGGAATGGGTCCACTCTACACACTCAGCCACATCGAGCGGAGCTG	2403							
QY	747	ctcgacgagatccacgacatgcccgaagcagggaaaccggtcgcgaagtcacccgatca	806							
Db	2402	GGGGCGAACAATGCCACCGCCGCTGGGTGAGCAGACCAAGTACAAAGCGGCTCCGACG	2343							
QY	807	acggtcggcgcaacgtacacctgttcgacacacccgcatgtggaataccgggacgctc	866							
Db	2342	CCGCTAGGGCGGAATGTGGCAGCTGTTCGATTCGCTCAGGTTGTGGGCTATGTCGGGCC	2283							
QY	867	c 867								
Db	2282	C 2282								

Search completed: October 6, 2002, 16:32:33
 Job time: 18386 sec

LOCUS	B96799	570 bp	DNA	linear	GSS 31-MAR-1998
DEFINITION	T32L9TR TAMU Arabidopsis thaliana genomic clone T32L9, DNA sequence.				
ACCESSION	B96799				
VERSION	B96799.1	GI:2998045			
KEYWORDS	GSS.				
SOURCE	the cross.				
ORGANISM	Arabidopsis thaliana Eunaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi				
REFERENCE	Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K., Berry,K., Granger,D., Sun,E., Wible,C., Adams,M.D. and Venter ,J.C. A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3				
JOURNAL	Unpublished (1997)				
COMMENT	Other_GSSs: T32L9TF Contact: Steve Rounsley Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: rounsley@igr.org Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 570.				
FEATURES	Location/Qualifiers				
Source	1..570 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="T32L9" /clone_lib="TAMU" /sex="hermaphrodite" /note="Vector: BelbacII; site_1: HindIII; site_2: HindIII ; Produced by Rod Wing"				
BASE COUNT	96 a 232 c 163 g 79 t				
ORIGIN					
Query Match	1.5%; Score 52.6; DB 12; Length 570;				
Best Local Similarity	45.5%; Pred. No. 1.9;				
Matches 187; Conservative	0; Mismatches 224; Indels 0; Gaps 0;				
Oy	578 gggaagcctatacaacctgctcgccgcggtcgagaaggcctatcgaaagtctcgggcg	637			
Db	4 gCTTCGGCGTTGCGCGAGCCGCCGCGGATACCCCGCGCTGGCGAGCACCTTGCTGA	63			
Oy	638 cgatcatcctaactcgggaaccggaatacaaaaagaacccgctcagcacgcgccatgagacct	697			
Db	64 CAATAACGCCCAAGGCCAACGGATTCTGGCGCCGCCGTGCTCTTCGCCGAGAGCCCGTGACC	123			
Oy	698 ctggggcccgcgaagacgcgctctacaagctgcgcgccttcgcacacaacacctcgaaagat	757			
Db	124 CTTCCTCGCGGCCCGGTGGCCCCGGAACGTCGTGAATCGAGAACAACCTTCACCGCGC	183			
Oy	758 ccagcgaactgcggagggcagggaaaccgcgctgcgaactacccgatcaacggtgcgcg	817			
Db	184 CGTGCTGCTGTCGGTGGGAGATCATCGGCGCCGACGCGCATCGATGCCGATGGCMC	243			
Oy	818 caaacgtaacctgtgtcgaacacacccgcgaatgtgggcataccgggcgcgttcggcaatcgt	877			
Db	244 CGGCTCTACAGCAGGCTGCTGCCATGGCTTCGGTCCGACACCCAGGCCGCCAAGACCGCCA	303			
Oy	878 gggcgccgcgctgcgcgaatggagacacacgctatctgagcacatcacactaactgaaga	937			
Db	304 GACCATCTACAGCGCCGATGCTGGCGCGACCGTCGGCGCATGTCTGAGCGCGACTTGATCT	363			
Oy	938 gaagctatcgccgacgaattcgccaacagggcccccttggttgaacgaact	988			

Db 364 CGAATACCTGCACCAACGTCGTGGGACACCGACCCAGCCGCGCAACATCT 414

RESULT 8
 LOCUS AM132314
 DEFINITION
 accession AM132314
 version AM132314
 keywords
 source soybean.
 organism Glycine max
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 448)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Coryell,V., Khanna
 'A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wille,T., Underwood,K., Stepien,M., Thelander,B., Allen,M., Bowers
 'Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harey,N., Schurk
 'R., Ritters,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 'R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Hunttsville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 Seq primer: -40RP from Gldco
 High quality sequence stop: 401.
 Location/Qualifiers
 1. 448
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl013-235"
 /clone_id="Gm-cl013"
 /tissue_type="Whole seedlings, 2-3 week old seedlings,
 greenhouse grown"
 /lab_host="XLI0-Gold"
 /note="Vector: pluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from whole seedlings of 2-3 week old greenhouse grown
 plants. The cDNA library was prepared using the Stratagene
 pluescript II XR cDNA library construction kit.
 Complementary DNA was synthesized from mRNA using a primer
 consisting of a poly (dT) sequence with a XhoI restriction
 site. EcoRI adapters were ligated to the blunt-ended cDNA
 fragments followed by XhoI digestion. The cDNA fragments
 were directionally cloned into the EcoRI-XhoI restriction
 site of the pluescript vector. The ligated cDNA fragments
 were transformed into XLI0-Gold host cells. This library
 was constructed by Dr. Randy Shoemaker and Dr. John
 Erpelidng."
 BASE COUNT 84 a 154 c 151 g 59 t
 ORIGIN
 Query Match 1.48; Score 51.4; DB 9; Length 448;
 Best Local Similarity 47.48; Pred. No. 3;
 Matches 154; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

[illegible]

RESULT	9
B1144217	
LOCUS	B1144217
DEFINITION	602907950F1 NCI_CGAP_Kid14 Mus musculus cdna clone IMAGE:5064900 5'', mRNA sequence.
ACCESSION	B1144217
VERSION	B1144217
KEYWORDS	B1144217.1 GI:14604218 EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1260)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

Tissue Procurement: Jeffrey E Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrived by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1175 row: c column: 13
High quality sequence stop: 405.

FEATURES	Location/Qualifiers
source	1. .1260

```

/organism="Mus musculus"
/strain="FVB/N"
/bb_xref="taxon:10090"
/clone_image="5064900"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SpOrf6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

```

Query Match	1.4%	Score 51.4	DB 10	Length 1260
Best Local Similarity	47.4%	Pred. No. 4		
Matches 154	0	Mismatches 171	Indels 0	Gaps 0

OY 571 ccgcgcgcgacgagcctatcaacctgtctgccgcgctgagcaaggccctatgcygaactc 630
 || | | | | | | | | | | | | | |
Db 628 cccnccccccacggggccccccccttcttcaccacccccacacccccccgacagctggccgcccc 687

Oy	631	tcggggygagatgcatctcctaacgagccggatcaacaagaacccgtcagcaccgccaatg	650
Db	668	ccggccgcacacggcgccggcccccgcgcgcacacacaggacaccccccttaccaccccccgcg	747
Oy	691	cgaacctcttgysgcccgcagacgagcgtctacagatcgtgcgcgcctctgcacacacccctg	750
Db	748	gcggccacacacacacccggcccccacagcccccccccgcgcgcgcgcgtacagccggcccg	807
Oy	751	acgagatcacagcagctgcgcgagggcagggaacccgcgtgcgaagtacccgatacaacgg	810
Db	808	ccacagcgccgcccccccccacgacacacacacccccgcacgcccgcgcacatcacagccccc	867
Oy	811	tcgggcgcgaacgtcacccctgtctgcacacacacccgcgcaltgttgagctaacgcgyscgctccgc	870
Db	868	cccccccgccgcgcgcgcgcgcctctcccccgcgcgcgcgcgccttcagacacccccgcgc	927
Oy	871	actcctgysgcgcgcgcgtgcgcga	895
Db	928	tcccgcgcgcgcgcgcgcgcgcga	952

RESULT 10
CNS01087
LOCUS CNS01087 961 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC
BA003H07 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098641
VERSION AL098641.1 GI:5610252
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 961)

TITLE	COMMENT
<p>Direct Submission Submitted (23-JUN-1999) Genoscope - Centre National de Séquençage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr)</p>	<p>Determination of this BAC end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'étude du polymorphisme humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.</p>

FEATURES	Location/Qualifiers
source	1. .961

```

/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_1id="DrosBAC"
/clone="BACN03H07"
/note="end : Sp6"

```

Query Match	1.48;	Score 51.2;	DB 12;	Length 961;
Best Local Similarity	25.58;	Pred. No. 4.1;		
Matches 69;	Conservative 91;	Mismatches 111;	Indels 0;	Gaps

QY	575	gcgagcagcgactatacaacctctctgccgcgtcagcaggcgactatycgacltctcgg	634
Db	678	ymgsasakaaaacwagmcmgmagsgscgcscscgscggcagagysagggngccscsgssssv	737

Oy 635 cggcgatgcatcctacgcggcacgcgatccacaagaaccggtcagacacgcgccatgcyac 694
 | :
Db 738 CGMAAAMAAA VMMAIRGVCAKCKASDSASVCMKGGCCGSCACCCCCSCSCSGC 797

[illegible]

RESULT	LOCUS	DEFINITION
11 CNS016KD/c		
CNS016KD	1100 bp	DNA linear GSS 26-JUL-1999
		Drosophila melanogaster genome survey sequence SP6 end of BAC
		BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION	AL10685
VERSION	AL10685.1
KEYWORDS	GI:5624152
SOURCE	GSS.
ORGANISM	fruit fly, <i>Drosophila melanogaster</i>

REFERENCE 1 (bases 1 to 1100)

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

COMMENT Determination of this BAC-end sequence was carried out as part of a

<http://www.eegp.ebi.ac.uk/>. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector paeloBAC11.

FEATURES	SOURCE	Location/Qualifiers
		1. .1100
		/organism="Drosophila melanogaster"
		/plasmid="pBelobAC11"
		/db_xref="taxon:7227"
		/clone_11b="DrosBAC"
		/clone="BACN16D22"
		/note="end : SP6"
BASE COUNT		132 a 229 c 106 g 220 t 413 others
ORIGIN		

Query Match	1.4%	Score 51.2;	DB 12;	Length 1100;
Best Local Similarity	24.2%	Pred. No. 4.2;		
Matches 146;	Conservative 184;	Mismatches 271;	Indels 3;	Gaps 2;

QY 506 gaaccgtgtcacgcaccacgcacatcgtctatgcttgaagaaacctgtgtctgac 565

Db 1078 G G M V V M G M V G S V G M R G M G M R V M M M V A M M M M M M R M V A R A R M M G M G V G Y M G M G 1019

[illegible][illegible]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

DD 338 MGRMRMGVGGVMG--GGRGMVGGVVMGVGGMGMGVGMVGMVGMGMGMGMGM 901

686 ccattgcgacctctgtgggccccgcagacgcgtctctacgagctgtgcgcctctgcacacac 745

Db 900 GMYGGMGVVZGMRRMGVGMMGRAGMCGMMRMVYMRMMMMMRGMGAMGSGCGG 841

QY	746	cccgagcagagatccgacgcacgcgcggcgaggcgaaacccggtctgcgaacgcaccacatc	807
Db	840	SMVSGCGCCGGCGMGRGVAGMGRSMRMAAGMGRGVGMGCACAGMGMGCACACACCGCGASM	78
QY	806	aacggtgcgcgcgaacgcttcacccctgtctgcacaccacccgcgaatgttggtgcataccggccgt	865
Db	760	MCASGSGSGGAMCAAGVGGMSGCGGCKRSGCCVCVCMCGGGAGCGGSGCCGCGGSGGCCGV	723

Db 720 CCGGCGSSCCYGGRCCTVTVSGGSGCSMSGCCMGCCSSSGGGCGSCCGCCCGGSC 661

Db 660 SGGSMCRCCCCSCCGASMGSCSCCGCGCCRCMCMMARACACAGCGGCCCCSACG 601

Db 600 CAGCAGCMSCCAGCAGCMGCGCCAMCMCMCCAGCGCAGCGAGCGAGCGAGTSSGVCW 5418

Db 540 GCGCGCGCMGYCRCGARCKCBGMSAGACGGDKGGRGAABAKAKAAGAARRRMAAGAAG 481

Db	480	GDBG	477
----	-----	------	-----

RESULT 12	
CNS006XK/C	
LOCUS	
CNS006XK	
935 bp	DNA
linear	CCS 03-TIN

LOCUS	DEFINITION	ACCESSION
CNS006XK	935 bp DNA	Linear GSS 03-JUN-2001
Drosophila melanogaster genome survey sequence T7 end of BAC # BAC1109 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		U066051

ORGANISM	<i>Drosophila melanogaster</i>
SOURCE	fruit fly.
KEYWORDS	GSS.
VERSION	GI:4945019
ACCESSION	AL066051
AL066051.1	

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 935)

NAME ENGLISH
 AUTHORS
 TITLE
 JOURNAL
 I (cases 1 to 555)
 Genoscope,
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
 BP 191 91006 EVRY cedex - FRANCE (E-mail : secret@genoscope.cnr.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP) for constructing a physical map of the Drosophila

Web : www.genoscope.cns.fr
E-mail : seget@genoscope.cns.fr

The BACs containing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see http://www.fruitfly.org/TheBACP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osogawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BGP from the isogenic strain y2; cn bw sp, the same strain used for the BDP.

pl and EST libraries. A more detailed description of the libraries and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES      location/qualifiers
              1. .935
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"

```

BASE COUNT	ORIGIN
257 a	170 c 162 g 96 t 250 others

Query March Similarity 1.4%; Score 51; DB 12; Length 935;
Best Local Similarity 30.5%; Pred No. 4.4;
Matches 137; Conservative 92; Mismatches 219; Indels 1; Gaps 1;

OY 563 gaccatgctcgagcggacgcagcatcatacaactgtctgccgcgttcgacgagccatg 622
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 907 GAGCCCGGSCGCCGCCGCCSSCSCSSSCSCSCSCSCSCSCSCSCSCSCSCSCC 848

OY 623 cgaagtctcgaggcgatgatcctaagcgacccgagatcaaaagaacctagcac 682
::
Db 847 SSGCGCCCGCSGCCGCCGCCS GCCSSSCGCCGCCGCCS CCGCGCCGCCGCCS GGC 788

OY 683 gccccatcgagcac -ctctggggcccccgagagcgctctacagcttggcgcccctcac 741
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 787 GSSGGCGGSGGCCGCCGCCSKKCGGCCGSGGSCSSSGSGGSGGSGGSCSSSC 728

OY 742 acaccctcgacgagataccacgactccggagagcaggaaaccgctcgcaactcaccc 801
:
Db 727 SCGGSSGSGSGGCCGCCGCCSCGCCSGCGSSGCGSGCGGCGCGCCGCCGCCSC 668

OY 802 gatcaagtgctggccgcgaacgttcacccctgtttcacacaccccgcatgtggcatccgg 861
:
Db 667 CCSCSSSSCGSSSCSSSCSSSCSSSCGCCGCCGCCGCCGCCGCCGCCGCCGCA 608

OY 862 ccgttcggcgaactccctggggcgccggttcgcgaattggagagacacgctatccgagaca 921
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 607 GCSCASGCGSSGAGATGASGASRSVSGSSSGASGCCCCGCGMCRAGSGKMGSNGSSG 548

OY 922 tccacctactcgacagagacgatcatcgccgcgaattcgccacagagccccctcgactga 981
:
Db 547 RCGGACSSGSGBSBSSKKRKGASCASCSAAGKSGSCGBKHGKACASSBTNTTTT 488

OY 982 agcaacttaagcaacttatctcgatcatt 1010
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 487 ATTAATAATATTTTTNTNMTATTATTAATT 459

RESULT 13
CONS0720/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACRI1B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL066742
ACCESSION AL066742.1 GI:4945205
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Neoptera; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Melzoza; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (O2-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
determination of this BAC-end and sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
Melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2: cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be

[illegible]

```
FEATURES
  source      Location/Qualifiers
              1..498
              /organism="Chlamydomonas reinhardtii"
              /strain="C9"
              /db_xref="taxon:3055"
              /clone="IC062e06_r"
              /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
              XhoI; The cDNA library was constructed from cells cultured
              in a carbon stress acclimatized condition in which carbon
              dioxide concentration in the bubbling gas was changed from
              5% to 0.04%"
BASE COUNT   90 a 184 c 146 g 78 t
ORIGIN

Query Match      1.4%; Score 50.4; DB 9; Length 498;
Best Local Similarity 46.9%; Pred. No. 4.9; Mismatches 216; Indels 6; Gaps 1;
Matches 196; Conservative 0;

QY 493 cctacgtgtccatgtaacgctgttcacgacacgacgacacatgctatgcttgaagaacc 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 CCCACAAGTGCAGAGAACATGTTCTCAGACAGAGCTGACCGTACGGCATGCCGATG 95

QY 553 ctgtgtgtctgacgatacgccgagcgagcgacgtctatcaacctgtccgccgctgcgacg 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 96 AGGTGCCCATCAGCCGATCGCGCCGCTCAAGGCGCATCTCCCTACGGCGCATCTAAGC 155

QY 613 agggcctatagcagcgttctcgagcgatgacatcctacgagcccgatatacaagaacc 672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 156 TGTTCAGAGAGGACATGTTCCGACCTGGCCGTGTCGCAAGAGACTGGCGATCCTGC 215

QY 673 cgttcacgacccgacatgacacctctggggcccgacgacgacgtctacgagctgcgcg 732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 TGTGTGGCTACTTCAACCCCATGGGCGGACCCCAAGCGGCGAGCTGGTGAGACACCCCG 275

QY 733 cctctgcaacacacctgacgagatcaacgacactgcgcgagcgaggaagaaacgcgctgcga 792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 TGGGATCTCCCAACACCTCATGCGTACATCCAGAGTGAGCTTGCGGCACAGC----- 330

QY 793 aggtacccgataaagtgctgcgcgcaacgtacacctgttctgacacaccccgatgtggg 852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 -CGAGTTCCTGCGTGTGTTGGCAACGACTACCCCAAGCCGCGAGCGACCGCATCGCG 389

QY 853 catcacggcgccgtcgacatcctggggcgccgctgcgcgaatggagacacgct 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 ACTACATCCAGCTCATGACCTGGCGAGGCCACGCTACGCTGTGCTAAGACCTT 447

RESULT 15
CNS01213      645 bp      DNA      linear      GSS 26-JUL-1999
LOCUS         Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION   BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION    AL101589
VERSION      AL101589.1 GI:5613200
KEYWORDS     GSS.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 645)
              Genoscope.
AUTHORS      Direct Submission
              Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the European Drosophila Genome Project (EDGP) -
              http://www.edgp.edi.ac.uk -. This Drosophila melanogaster BAC
              library (Dros BAC) was made by Alain Billand at CEPH (Centre
```

```

d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
  source      Location/Qualifiers
              1..645
              /organism="Drosophila melanogaster"
              /plasmid="pBelOBAC11"
              /db_xref="taxon:7227"
              /clone_lib="DrosBAC"
              /clone="BACN08C07"
              /note="end : 17"
BASE COUNT   28 a 26 c 85 g 92 t 414 others
ORIGIN

Query Match      1.4%; Score 50; DB 12; Length 645;
Best Local Similarity 9.5%; Pred. No. 6.3; Mismatches 249; Indels 0; Gaps 0;
Matches 40; Conservative 131;

QY 293 gtccctgagaagctggctgcacgacgacgctggcgtcacgacgagagctggggcgtga 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 GTTTNNNGNNNNNSCCSCNCTNCTNNNTTTNSSSSSSNNNTNNNNNSNNNSNN 236

QY 353 ccggacgctgactgcgacgagggcgctgagctgcttaccatcgaaagcgaaccgctgtgt 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 NNSSSGSSNNNNNSGSSNNNSNNNSGSSSSNSTSSNNNTNNNNNSNNNSNN 296

QY 413 catcgagctctgtgtacacccatgagatgcttggatgctgactggcgccgacacct 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 TNSNNNSNNSSSTNTNTNSTSSNNNSATSSSSNSSSGTTBSGSSSSSSNNNN 356

QY 473 cgctggcgctgcttaccgctcctacgctgctacgacgctgctacgacacccgacacat 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 NNASNNBSBNSNNNSNNTSTTNGSSGSSSSSSSTSSNTSTSTSSNSNNSTT 416

QY 533 cgtctatgcttgaagaacctgtgtctgacgagatgcccgcgcgagcagcgtatcaa 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 SSSSTTSTSTSSSSNSTSSSTTTTNSASNNNSNNNNNGSSSSNANSTTNS 476

QY 593 cctgctgcggcgctgcgacgagggcctatgacagcttctcgcgcgatgcatcctacgg 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 NTTTTSSNNSSSCTNTSANTSTNTTNTAGNNNATSSSNNNSNNNSNNNSGSG 536

QY 653 gcaaccgatacaaaagacccgctcagcagccacatgacacctctggggcccccgcaga 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 TSSGNTSSSCTGTBSGNSNNNSNNNSCNSNNNSCTGSSSTTTGSSSSSSNSAA 596
```

Search completed: October 6, 2002, 16:22:38
Job time: 17806 sec

DY 361 tgaatcgagaaaggcgccttgagctgcttacaatcaagaacccgttgtaactgcaat 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2792 AAGACCCGCGGTGGGCCCTTAGCGCCCGGGTAATCATGAGCGAACCACCAACGCCTGGACA 2733
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 ccttgatcatcaacgatcgagatgcttcgtagctgtagctggcgcaagacctgctggc 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2732 ACCTGAGTGTCTGGAGAGTGAACCAATTCAGAGCAGACGCTCCGAGCGCTCAAGCGCCGG 2673
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 481 tgcctcaacgcttctaagtgatg-----aacggtcaacacaccggaataatcg 534
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2672 GGTCCATTCCGCTGCCCCAACGCGATCGTGCGCAATCGCCCACACGCGCACGCGAG 2613
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 535 tcatactctgaagaacctgtgtctgaccgagtccgcgcygagcgactaacaac 594
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2612 TGTGGGACTCAACGCGCCCTGTTCCAGCGACCGAAATACGCGGGGCTAAGCGCTCGAT 2553
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 595 tgctgcgcgcgtcgagcagagcctatgctgagctgtctggcgcgatgaatcactaagc 654
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2552 ACAATGGCGGGGTGCGCCGAAGCGCTTTCGGCGCGCGCTGCAGCGGACGAGTTACTAG 2493
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 655 accggaatcaaaaagaaccgcgtcagaacccgccatgycgaaccttggygcccccgagcg 714
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2492 GCCATATGACCAAACCCCGGCAATGCGCTGGGAAAAGGAAATGGCTCCACTCAATC 2433
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 715 cgctctagagctgcygcgcctctgcacacacctcgacagatccaagcagctgcccag 774
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2432 TCTACACATCATCACACACATGAGAGCGGAGCTGGCGCGCAACATGCGACGCGCGTGC 2373
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 775 caaggaaaccgcgtcgcaacgtcaaccgatacaagcgctggcgccgaagctcaacctgtcg 834
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2372 GTACGACGACACGATCAAAAGCGGCTCCAGACCGCGCTAGAGCGGGAATTGCGACATGTCG 2313
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 835 acaccaccgcgatgtggcattaccgscgcgtcc 867
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2312 ATTCCGTCAAGTTGTGGGCTATGTCGCCGCC 2280
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-07-977-630-2
Sequence 2, Application US/07977630
Patent No. 5583038
GENERAL INFORMATION:
APPLICANT: Stover, Charles K.
TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Baln, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,630
FILING DATE: NO. 5583038ember 17, 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Herion, Charles J.
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 469201-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 6407 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: nucleic acid
; US-07-977-630-2

```

Query Match	2.38;	Score 82.2;	DB 1;	Length 6407;
Best Local Similarity	47.88;	Pred. No. 2.1e-10;		
Matches 274; Conservative	0;	Mismatches 293;	Indels 6;	Gaps 1;

QY	301	aaagctggtgcgaacgaagccgctgtgcgttcagccgagagactgtgtgcgttcacgcaagc	360
Db	3557	AGCTCTB6GCTGCCGTACTGCGCCGCTG6CAAGGACGATGTGCTCAAGG6GATCTACGCC	3616
QY	361	tgactcgcagagagcgctgtgagctgtgcctctacatcgaaagacccgttggltcalgtcagt	420
Db	3617	AAACCCGCGCGTGGGCCCTAGCGCCGCGGATCATGAGGCGAACCACAGGGGCTGGCA	3676
QY	421	ccttggttcatacccgatcgatgatagtcttgatgtctgacttggcgcgagactcgtctggc	480
Db	3677	ACCTGCGTGGTGTGGAGCTAGACATCCAGACGCGCTCCGACGCTCAAGCGCCCGGG	3736
QY	481	tgccctcacgctcactcagtggtcagt-----aacgtgttcagcacaacggagacatcg	534
Db	3737	GGTCCATCCGCTGCGCCACAGCGATGTGGGCAATGCGCCACAGCGCCACAGCACGACG	3796
QY	535	tctatgcttgaagaaacctgtgtgtcttcacgatagcgcgcgcgcgcactcaaac	594
Db	3797	TGTGGGCATCAACGGCCCGTTCACAGGACCGAATATACGGCGGGGCTAAGCGCGTCCGAT	3856
QY	595	tgctgcgcgcgttcgagcagcgccctatgtcgcagcttctcgcgcgcgtatcactcaagcgc	654
Db	3857	ACAAGGGGGGTGCGCGGAAGGGCTTTCGGCGCGGTGCACGCGCACGAGTTACTACG	3916
QY	655	accggtacataaagaagaaacctcagaacgcgcacatgcgcacctctggcgcccgcaagcg	714
Db	3917	GCCCTATGACCAAAAACCCGGCCACATGCGCTGGGAAACGGAAATGGCTCACTATGATGC	3976
QY	715	cgctctacgagctgcgcgcctctgcacacacacctcgcagagatltcaacgcactgtccgaag	774
Db	3977	TCTACACATCAACGCAACATGGAAGCGGACCTGGGGCGAACAATGCCACGGCGGCTGGC	4036
QY	775	caaggaaaccgcgttcgcaacgtlcaaccgcatacgcgltcggccgcgaacgltcaacctgtcg	834
Db	4037	GTCAGCGACACAGTACCAAGCGGCTCCAGCCCGCTAGGGGGGAATTCGCACATGTCG	4096
QY	835	aacacaccgcacgttgggatacgcgcgttc	867
Db	4097	ATTTCGTGACGTTGTGGGCTTATCGTCCGGCC	4129

RESULT 3
 US-07-977-630-20
 : Sequence 20, Application US/07977630
 : Patent No. 5583038
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Stover, Charles K.
 :
 : TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING
 : TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS
 :
 : NUMBER OF SEQUENCES: 84
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Carella, Byrne, Bain, Gillfillan, Cecchi,
 : ADDRESSEE: Stewart & Olstein
 : STREET: 6 Becker Farm Road
 : CITY: Roseland
 : STATE: New Jersey
 : COUNTRY: USA
 :
 : ZIP: 07068
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: 3.5 inch diskette
 :
 : COMPUTER: IBM

```

OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-07/977,630
FILING DATE: NO. 5583038ember 17, 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heron, Charles J.
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 469201-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4120 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: nucleic acid
US-07-977-630-20

```

Query Match	2.3%	Score 80.6	DB 1	Length 4120
Best Local Similarity	47.6%	Pred. No. 4.4e-10		
Matches 273	Conservative 0	Mismatches 294	Indels 6	Gaps 1

QY	301	agaagctgctgcacacgaagaacgcgtctgagctcgaacgaagaagctctggcggtctacggaacg	360
Db	2382	AGCTCTGGCTCCCTACTAGGCCCCGTGGCAAGCAGCAGCATCTGCTCGAGGGGATCTACGGCC	244111
QY	361	tgactcgcgaagaggcgctlgaagctlgtcctacatcgaagcgaaccgctlgtcalgtcagt	420
Db	2442	AAACCGCCGGGTGGCCCTTAGGCGCCGGAGTATCTTAGGCGCAACCAACAGCGGCTGGCA	2501
QY	421	ccttggtcalcacccgatcggatgtgtcttggaagtgtcgtactggcgcgagaactcgtctggc	480
Db	2502	ACCTGCTGGTGTGGAGGTGAGCATTCAGAACCAACGACGCTCCGAGAGCGCTCACGCCCCGG	2561
QY	481	tgcttcacgcgtctcactagctgtctcagt-----aacggtltaacgaacgaagacacatcg	534
Db	2562	GGTCCCATTCGGCTCTGCCCCAACGCGATGTGGGCAATTGGCCCAACGGCCACAGCAGCGAG	2621
QY	535	lctatgctctlgaagaagaacctgtgtctctacagatgcgcgcgcgcgcgcgcgcgcctatcaac	594
Db	2622	TGTGGGCACTCAAGGCCCTGTTCACAGCAGCAATAGCGGGGGGTAAAGCCGCTCGCAT	2681
QY	595	tgctcgcgcgcgtcgcgaagagccctatgcgaagctlctcgcgcgatgcattcctacgggc	654
Db	2682	ACATGGGGGGGTGGCGGAAGGCGCTTCGGCGCGCGCTGCATGGGACGCAAGTACTCAG	2741
QY	655	accggtatcaacaagaagaaccgcgtcagaacgcgcctatgtcgcacctctcggggccccgaagcg	714
Db	2742	GCCCTCATAGCAAAAAMACCCCGGCACATAGCCTCGGGAAACGGAAATGGCTCCACTGATGC	2801
QY	715	cgctctacgaagctlbgcgccctcgaacaacacctcgaagagatccaagcactgcgcgagg	774
Db	2802	TCTACACATCTAGGCACATGAGAGGCCGAGCTGGCGCCGAACATGCCCACCGCCGCTTGGC	2861
QY	775	caaggaaaccgcgctgcgaacgltcaaccgatcaacggtlbgcgccgcgaacgltcaacctgttcg	834
Db	2862	GTCAGCAGACCACTGACAAAGAGGGCTTCGACGCGCGCTGAGGGGGAATTCGCACTGTTCG	2921
QY	835	acacaccaccgcatgttggcattaccgagcgctgc	867
Db	2922	ATTCCGTCAGGTTGTGGCTATCTGTCGCCCC	2954

RESULT 4
US-07-977-630-21/c
; Sequence 21, Application US/07977630
; Patent No. 5583038
; GENERAL INFORMATION:

```

1  APPLICANT: Stover, Charles K.
2  TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING
3  TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS
4  NUMBER OF SEQUENCES: 84
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Carella, Byrne, Bain, Gillfillan, Cecchi,
7  ADDRESSEE: Stewart & Olstein
8  STREET: 6 Becker Farm Road
9  CITY: Roseland
10 STATE: New Jersey
11 COUNTRY: USA
12 ZIP: 07068
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: 3.5 inch diskette
15 COMPUTER: IBM
16 OPERATING SYSTEM: MS-DOS
17 SOFTWARE: ASCII
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/07/977,630
20 FILING DATE: No. 5583038ember 17, 1993
21 CLASSIFICATION: A35
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Herion, Charles J.
24 REGISTRATION NUMBER: 28,019
25 REFERENCE/DOCKET NUMBER: 469201-174
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 201-994-1700
28 TELEFAX: 201-994-1744
29 INFORMATION FOR SEQ ID NO: 21:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 4120 nucleotides
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: circular
35 MOLECULE TYPE: nucleic acid
36 US-07-977-630-21

```

Query Match	2.3%	Score 80.6	DB 1	Length 4120
Best Local Similarity	47.6%	Pred. No. 4.4e-10		
Matches 273, Conservative	0	Mismatches 294	Indels 6	Gaps 1

OY	301	agagctgtctgcccgaagaacccgtctgctgcacccgaagaagctctgggagctaccgagc	360
Db	1739	AGCCTGTGCTCCCTACTGTGGCCCTGGCAAGCAGCATCTGCTCCAGGGGAGCTACCC	1680
OY	361	tgactcggcagagggcgctgtagctgccttacatcgaaagcgaaccgcttgytcatgca	420
Db	1679	AAACCCCGCGGTGCGCCCTTAGGCGCGCGGTACATCGAGGCGAAACCCACAGCGCTGGCAA	1620
OY	421	ccttgagcatcacacgcgatcgatgcttgatgtgtacgtggcgcgagacctcgttgcc	480
Db	1619	ACCTGTGTGTGTGGAGGTAGACATTCACAAACGACGCTCCGAGGGCTCAAGCGCCGG	1560
OY	481	tgcttcaacgcttctacgtgtgcatg-----aacggttcacgaacacgcgaacatcg	534
Db	1559	GGTCCATCGCGTGGCCCAAGCGCATGTGGGGCATGTGGCCCAAGCGCACAGCACAGAG	1500
OY	535	tctatgcttgaagaacacctgtgtgtcttcacgaatgcgcggtgcgaacgactataacc	594
Db	1499	TGTGGGACATCAAGCGCCCTGTTCACAGCACCCCAATATCCGGGGTAAAGCGCTTCGAT	1440
OY	595	tgctcgcccgctctgagcagaagcctatgcagcttctcggcgagatgatcctacggc	654
Db	1439	ACATGGGGGCTGGCGCGAAGGCCCTTCGGGGCGCTCCATATGGCCACACGCAATTACTAG	1380
OY	655	accgagatcaaaagaacccgcctcagaacccgcacatgcgaacctctgysgcccgaagcg	714
Db	1379	GCTCATATGACCAAAAAACCCCGGCACATGCTCTGGAAACGGAATGGCTCCACTCATATC	1320
OY	715	cgctctacgagctcgtgcgacctctgcacacacaccttgaagagatctcaagatctgcggag	774
Db	1319	TCCTACACTATGACCATATGAGGCGAGGCTGGCGCGAATGTCACACGCGCGCTTGC	1260

QY	835	acacacacccgcatgtggtacatccgagcgttc	867
Db	1199	ATTCCGTCAGAGTTTGGGCTATCGTCCGCC	1167

RESULT 5
US-08-316-950-11

```

Sequence 11, Application US/08316950
Patent No. 5679515
GENERAL INFORMATION:
APPLICANT: STOVER, Charles K.
APPLICANT: HICKEY, Mark J.
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
TITLE OF INVENTION: thereof
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,950
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 15371A-001100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 4120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (plasmid)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..4120
OTHER INFORMATION: /standard_name= "plasmid pMV206"
US-08-316-950-11

```

Query Match	2.3%	Score 80.6;	DB 1;	Length 4120;
Best Local Similarity	47.6%	Pred. No. 4.4e-10;		
Matches 273; Conservative	0;	Mismatches 294;	Indels 6;	Gaps 14.

QY	301	agaagctgctgcacgaacgaacgcctgctgcacacggagaagctctggggtgtacgcgacg	360
Db	2382	AGCTCTGCTGCTCCGCTACTGGCCGCTGTGGACAGCAGCATGTGTCTCGAGGGGATCTACCC	2441
QY	361	tgactcgcgacgagggcgctgtagctgctctaatctgaacgcgaaccgctgtgcatcagct	420
Db	2442	AAAGCCGCGGCTGGGCCCTTAGCGCCGCTACTCTGAGGCGAACCACACAGCGCTGGCA	2501
QY	421	ccttgatcatcacagatctcgaatgcttcgtagtgcctacgtgcgcgacagccctgcgtggc	480
Db	2502	ACCTGCTGTGCTGTGGAGAGTATGACCAATCCAGACGACGCTCCGAGCGCTCAACGCGCCGG	2561

QY	481	tgcttcacgcgtctctagatgtcatg-----aacggttaccgcaacacggaacatcg	534
Db	2562	GGTCCCATCCGGCTGCCCAACGCATGTGTGGCAATGTGGCCAAAGGCCAACGCACACGCAG	2621
QY	535	tctatgccttgaagaaacctgtgtgtctaccgagtgcgagcgcgagcgcctalcaacc	594
Db	2622	TGTTGGGCATCAACAGGCCCTGTGTTCCACGCACGCAAAATACGCGGGGGTAAAGCCGCTCGCAT	2681
QY	595	tgctcgccgcgcgttcgagcagggacctatgtcgaaagttctcgcgcgagatgaactcctagcgc	654
Db	2682	ACATAGGGGGGTGCGCGCAAGGGCTTTGGCGCGCGCTCATGTGGCACGCAGATTACTAG	2741
QY	655	acggatcaccaagaagaacctcgtacgagcgcgcacatgtcgacctctctgtgggcccccgagagc	714
Db	2742	GCTCTATGACCAAAAACCCCGGCGCACATGCTCTGGGAAAGGAATGCTTCACCTAGATC	2801
QY	715	cgtctctacgagctgtgcgcgcctctgcacacacctctgacgagatltcacgcaatgccgagag	774
Db	2802	TCTTACACATCTACGACCAATAGAGGCGGAGCTGGCGCGCAACATGCGACACCGCGCGCTGGC	2861
QY	775	caaggaaacctgcgtcgaacctcaacctgaatcgaatgcgtgcgcgcaagtcacactgtctgc	834
Db	2862	GTCAGCGAGACCAAGTACAAAAGCGGCTTCCACACGCGCTGAGGGGGAATGCGACATGTTGCG	2921
QY	835	acaccacccgcattgtggagctaacgagcgacgtcc	867
Db	2922	ATTTCGTGATGTTGTGTGGCTATGCTGCCGCCC	2954

RESULT 6
PCT-US95-12642-11

```

1 Sequence 11, Application PC/70US9512642
2  GENERAL INFORMATION:
3  APPLICANT: Pathogenetics Corporation
4  APPLICANT: 201 Elliott Avenue West, Suite 150
5  APPLICANT: Seattle, Washington 98119
6  TITLE OF INVENTION: Mycodacterial Reporter Strains and Uses
7  TITLE OF INVENTION: Theoreof
8  NUMBER OF SEQUENCES: 18
9  CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Townsend and Townsend Kourile and Crew
11 STREET: Steuart Street Tower, One Market Plaza
12 CITY: San Francisco
13 STATE: California
14 COUNTRY: US
15 ZIP: 94105-1493
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentin Release #1.0, Version #1.25
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: PCT/US95/12642
23 FILING DATE: 02-OCT-1995
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/316,950
27 FILING DATE: 03-OCT-1994
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Hunter, Tom
30 REGISTRATION NUMBER: 38,498
31 REFERENCE/DOCKET NUMBER: 15371A-001100
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (415) 543-9600
34 TELEFAX: (415) 543-5043
35 INFORMATION FOR SEQ ID NO: 11:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 4120 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41 MOLECULE TYPE: DNA (plasmid)
42 FEATURE:

```



```

: GENERAL INFORMATION:
: APPLICANT: Pathogenesis Corporation
: APPLICANT: 201 Elliott Avenue West, Suite 150
: APPLICANT: Seattle, Washington 98119
: TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
: TITLE OF INVENTION: thereof
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend Kourile and Crew
: STREET: Steuart Street Tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/12642
: FILING DATE: 02-OCT-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/316,950
: FILING DATE: 03-OCT-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunter, Tom
: REGISTRATION NUMBER: 38,498
: REFERENCE/DOCKET NUMBER: 15371A-001100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4296 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (plasmid)
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..4296
: OTHER INFORMATION: /standard_name="plasmid pmh28"
: PCT-US95-12642-15

Query Match      2.3%; Score 80.6; DB 5; Length 4296;
Best Local Similarity 47.6%; Pred. No. 4.5e-10;
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

QY 301 agagctgctccacgcaagccgctgagcgtacagccgagaagctctgggagctacggagcagc 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2405 AGCTCTGCTGCTCCGACTGCGCCGCTGCGCAACGACGATCTCTGAGGGGATCTACCCCC 2464

QY 361 tgaactcgagagggcgctgagctgacctacatcagaacgagccgttggatcagtcagt 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2465 AAAGCCGGCGCTGAGCCCTAGAGCGCGGTACATCGAGGCGCAACCCACAGCGGCTGGCAA 2524

QY 421 ccttgatcatcagcatcgaatgcttgcgagtgtgactgagcgagccgagacctgctggc 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2525 ACCTGCTGCTGCTGAGCTAGCATCCATCCAGCAGCGCTCCGAGCGCTCAGCCCGCGG 2584

QY 481 tgccttaacgcttctacgtgtgtcattg-----aacggtgtacgacacacgagacacatcg 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2585 GGTCCCATCCCGCTGCGCCACCGCATCGTGGCAATCGCGCCACACGCGCACACGCGAG 2644

QY 535 tcatatgcttgaagaacctgtgtgtctgacgagatgcccgcggcgagcgactacaacc 594
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2645 TGTGAGGACTCAAGCGCCGCTTCCACGACGCAATACGCGCGGGGTAAAGCGGCTCGCAT 2704

QY 595 tgcctgccgcgctgagagagccttatgtgcagcgttcttcggcgcgagatgcatcctcagggc 654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 2705 ACATGCGCGGCTGCGCCGCAAGGCCCTTTCGGCGCGCGCTGATGCGCAGCCAGTACTCAG 2764
QY 655 accggtatacaagaagaccgctcagcagccgcatgagccctctggggcccccagagacg 714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2765 GCCTATATACCAAAAACCCCGGCCATCTGCTGGGAACGGAATGGCTCCACTCAGATC 2824

QY 715 cgcctacgagctgcgcgcctcgacacacacctgacagagatccacgcaactgcccgaag 774
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2825 TCTACACACTGACGACATCGAGCGGAGCTGCGCGGCAACATGCGACCGCGCGCTGGC 2884

QY 775 cagggaaccgcgctgcgaacgttaaccgataagcgttgcgcgcaagctacacctgttcg 834
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2885 GTCACGACGACCATGACAAACCGGCTCCGACGCGCTGAGGCGGAATTGCGCACTGTTCCG 2944

QY 835 aacaccaccgcatgtggcagatccggcgctcc 867
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2945 ATTCCGTGAGGTGTGGGCTATGCTCCCGCC 2977

RESULT
9
US-08-316-950-16
: Sequence 16, Application US/08316950
: Patent No. 5679515
: GENERAL INFORMATION:
: APPLICANT: STOVER, Charles K.
: APPLICANT: HICKEY, Mark J.
: TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
: TITLE OF INVENTION: thereof
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend Kourile and Crew
: STREET: Steuart Street Tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/316,950
: FILING DATE: 03-OCT-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunter, Tom
: REGISTRATION NUMBER: 38,498
: REFERENCE/DOCKET NUMBER: 15371A-001100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4352 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (plasmid)
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..4352
: OTHER INFORMATION: /standard_name="plasmid pmh29"
: US-08-316-950-16

Query Match      2.3%; Score 80.6; DB 1; Length 4352;
Best Local Similarity 47.6%; Pred. No. 4.5e-10;
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

QY 301 agagctgctccacgcaagccgctgagcgtacagccgagaagctctgggagctacggagcagc 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2405 AGCTCTGCTGCTCCGACTGCGCCGCTGCGCAACGACGATCTCTGAGGGGATCTACCCGCC 2464
```

361 tgaactgcagagggcgctgagctgcttaccatogaagaccggttgctatgcagt 420
2465 AAGAGCGCGCGCTGCGGCCCTAGAGCGCGGATCATCGAGCGAACCAGACGCGTGGCAA 2524
421 ccttggtacacacgatgagatgcttggaatgctgagctgagcgccgagacccctgcctggcg 480
2525 ACCTGCTGGTGTGAGCGTAGACCATTCACAGCAGCGCTCCGAGCGCTCAGCGCGCCGGG 2584
481 tgccttaccgctcctcagtgctcatg-----aacggttcaagacacccggacacatcg 534
2585 GGTCCCATTCGCTGCCAACGGGATGTGGGCAATGCGCCCAAGCGCCACGACGAG 2644
535 tctatgcttgaagaacctgtgtctcagatgcgagcgcgagcgagcctataacc 594
2645 TGTGGGCACTCAACCGCCCTGTCCACGACCGCAATACCGCGCGCTAGACGCGCTGCAT 2704
595 tgcctgcgcgctgagcgagcgctatgagagcttctgcgagcgatgataccttaagcg 654
2705 ACATGGCGGCGTGCAGGAGGCTTGGGCGCGCGCTCATGAGCGACGATTACTCAG 2764
655 accggtacacaaagaacccgctcagaccccgatcgagcccttgaggcccgagcg 714
2765 GCCTCATGACCAAAAACCCCGGCGCATGCGCTGGAAACGGAATGCTCCACTCAGATC 2824
715 cgcctcagagctgagcgccctcgcacacacccctcagagatcagcagactgcggaag 774
2825 TCTACACACTCAGCCATCGAGCGCGAGCTCGGCGCAATGCGCGCGCGCTGCG 2884
775 cagggaacccgctgcgaacgctcagacccgatcaagctgcgagcgagcagcttcttg 834
2885 GTACAGACACACGACGTAAGGCGGCTCCAGCGCGCTAGGCGGAATTGCGGCACTGTTCG 2944
835 acacaccccgatgtgagcataccggcgctcc 867
2945 ATTCCGTACAGTTGTGGCGCTATCGTCCGCC 2977

RESULT 10
PCT-US95-12642-16
Sequence 16, Application PC/7US9512642
GENERAL INFORMATION:
APPLICANT: Pathogenesis Corporation
APPLICANT: 201 Elliott Avenue West, Suite 150
APPLICANT: Seattle, Washington 98119
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12642
FILING DATE: 02-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,950
FILING DATE: 03-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 15371A-001100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4352 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (plasmid)
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..4352
OTHER INFORMATION: /standard_name="plasmid pmh29"
PCT-US95-12642-16

Query Match 2.3%; Score 80.6; DB 5; Length 4352;
Best Local Similarity 47.6%; Pred No. 4.5e-10;
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

301 agagctgctgctcagcagcagcgctgctcagcagagagcttggtgctacggcagc 360
2405 AGCTGTGCTGCGCGTACGCTGCGCGCTGCAAGCAGCATGCTCGAGGGAGTCTACCGCC 2464
361 tgaactgcagagggcgctgagctgcttaccatogaagaccggttgctatgcagt 420
2465 AAGAGCGCGCGCTGCGGCCCTAGAGCGCGGATCATCGAGCGAACCAGACGCGTGGCAA 2524
421 ccttggtacacacgatgagatgcttggaatgctgagctgagcgccgagacccctgcctggcg 480
2525 ACCTGCTGGTGTGAGCGTAGACCATTCACAGCAGCGCTCCGAGCGCTCAGCGCGCCGGG 2584
481 tgccttaccgctcctcagtgctcatg-----aacggttcaagacacccggacacatcg 534
2585 GGTCCCATTCGCTGCCAACGGGATGTGGGCAATGCGCCCAAGCGCCACGACGAG 2644
535 tctatgcttgaagaacctgtgtctcagatgcgagcgcgagcgagcctataacc 594
2645 TGTGGGCACTCAACCGCCCTGTCCACGACCGCAATACCGCGCGCTAGACGCGCTGCAT 2704
595 tgcctgcgcgctgagcgagcgctatgagagcttctgcgagcgatgataccttaagcg 654
2705 ACATGGCGGCGTGCAGGAGGCTTGGGCGCGCGCTCATGAGCGACGATTACTCAG 2764
655 accggtacacaaagaacccgctcagaccccgatcgagcccttgaggcccgagcg 714
2765 GCCTCATGACCAAAAACCCCGGCGCATGCGCTGGAAACGGAATGCTCCACTCAGATC 2824
715 cgcctcagagctgagcgccctcgcacacacccctcagagatcagcagactgcggaag 774
2825 TCTACACACTCAGCCATCGAGCGCGAGCTCGGCGCAATGCGCGCGCGCTGCG 2884
775 cagggaacccgctgcgaacgctcagacccgatcaagctgcgagcgagcagcttcttg 834
2885 GTACAGACACACGACGTAAGGCGGCTCCAGCGCGCTAGGCGGAATTGCGGCACTGTTCG 2944
835 acacaccccgatgtgagcataccggcgctcc 867
2945 ATTCCGTACAGTTGTGGCGCTATCGTCCGCC 2977

RESULT 11
US-08-316-950-13
Sequence 13, Application US/08316950
Patent No. 5679515
GENERAL INFORMATION:
APPLICANT: STOVER, Charles K.
APPLICANT: HICKEY, Mark J.
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Stewart Street Tower, One Market Plaza

```

1      CITY:  San Francisco
2      STATE:  California
3      COUNTRY:  US
4      ZIP:  94105-1493
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE:  floppy disk
8      COMPUTER:  IBM PC compatible
9      OPERATING SYSTEM:  PC-DOS/MS-DOS
10     SOFTWARE:  Patentm Release #1.0, Version #1.25
11     CURRENT APPLICATION DATA:
12     APPLICATION NUMBER:  US/08/316,950
13     FILING DATE:  03-OCT-1994
14     CLASSIFICATION:  435
15
16     ATTORNEY/AGENT INFORMATION:
17     NAME:  Hunter, Tom
18     REGISTRATION NUMBER:  38,498
19     REFERENCE/DOCKET NUMBER:  15371A-001100
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE:  (415) 543-9600
22     TELEFAX:  (415) 543-5043
23     INFORMATION FOR SEQ ID NO: 13:
24
25     SEQUENCE CHARACTERISTICS:
26     LENGTH: 4505 base pairs
27     TYPE: nucleic acid
28     STRANDEDNESS: single
29     TOPOLOGY: linear
30     MOLECULE TYPE: DNA (plasmid)
31     FEATURE:
32     NAME/KEY:  misc feature
33     LOCATION:  1..4505
34     OTHER INFORMATION:  /standard_name= "plasmid pMW261"
35
36     US-08-316-950-13

```

Query Match 2.38; Score 80.6; DB 1; Length 4505;

QY	301	agaagctgctcccaagcaagaagccctggggtctcaagccgagaagctctggggtctaccggcgaag	360
Db	2382	AGCTCTGGCTGCTCCGCTACTGGCCCGCTGGCAAGCGACGATCTGCTCGAGGGGATCTACCGCC	2441
QY	361	tgactcggcagaaggcgtctgagctgtgctctacatcgaagcgaaaccggttgatctgca	420
Db	2442	AAAGCGCGGCTGGCGCCCTAGGACCCCGGTACATGAGAGGCAACCAACGCGCTGGCAA	2501
QY	421	ccttggtatcaaccgatacgatgcttgatgctgacatggcccgacagccctgcgtgggc	480
Db	2502	ACCTGCTGTGTGGACCTAGACATCCAGACGCAAGCGCTCCGAGCGCTCAAGCGCCGGG	2561
QY	481	tgctctcaegtlcctacgtgtctcaatg-----aaccgltgcagcaaccgagacacatcg	534
Db	2562	GGTCCATCCCTGGCCCAACGATGTTGGGCAATCGCGCCAAACGCGCCACGCAACGCGAG	2621
QY	535	tctatgcttgaagaagacctgtgtgtgcacgaatggccggcgaggagcaagcctatcaaac	594
Db	2622	TGTGGGCACTACAGCCCTCTGTTCCACGCCACCGAATACGGCGGGGTAAAGCGCTGCAT	2681
QY	595	tgctcgcccgctgcgagcaaggtcctatgtcgagctgtctcggcgcgatgatctctacgggc	654
Db	2682	ACATGGCGGCGTGGCGCCGAAAGGCTTTGGCGGCGCTGGATGGCACCGCATTTACTGAG	2741
QY	655	acggatctacaagaagaccggtctcagcacgcgcctcgtgagcctctcggggcccccgaag	714
Db	2742	GCCCTATGACCAAAAACCCCGGCGACATGCGCTTGGAAAACGGAATGGCTCCACTCAGATC	2801
QY	715	cgctctaaagctgcgcgcctcgcagacacacctcgacagagatctcaacgcatgcgcgag	774
Db	2802	TCTACACACCTACGACCACTACGAGCGCGAAGCTGGGGGCAACATGGCAACGCGCGCGTGGC	2861
QY	775	caaggaaaccgcgctctgcgaacgtacacgcgatacaaggttcggccgcaacgctacccgtgtcg	834
Db	2862	GTCAGCAACACCAAGTAAACGAGGCTTCGACGCCCTCTAGGGGGGATATCGCATGTTTG	2921

Oy 835 acaccacccgcgatgtggcataccggygccgtcc 867
| | | | | | | | | |
Db 2922 ATTCGTCAGTGTGGGCCTATCGTCCGCC 2954

```

RESULT 12
PCT-US95-12642-13
Sequence 13: Application PC/TUS9512642
GENERAL INFORMATION:
APPLICANT: Pathogenesis Corporation
APPLICANT: 201 Elliott Avenue West, Suite 150
APPLICANT: Seattle, Washington 98119
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
TITLE OF INVENTION: Theroof
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12642
FILING DATE: 02-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,950
FILING DATE: 03-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 15371A-001100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 4505 base pairs
Type: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: DNA (plasmid)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..4505
OTHER INFORMATION: /standard_name="plasmid pmw261"
PCT-US95-12642-13

```

Query Match	2.38;	Score 80.6;	DB 5;	Length 4505;
-------------	-------	-------------	-------	--------------

Qy	301	aaagccttgagctgcgcacgaagaacccctctgacgctcaagcccgagagctctggggtatccggcagc	360
Db	2382	AGCTCTGGCTGCGGCTACTGGCCCTCTGGCAAGCAGCCATCTGTCGAGGGGATCTACGGCC	2443
Qy	361	tgactcgcagagagcgccttgagctgcctctacatcgaacgcgaaccggtctgctatgcagct	420
Db	2442	AAAGCCGCGGGTGGGCGCTTAGCGCGCGGTACTCTGAGGCGAACCACACAGCGCTGGCA	2503
Qy	421	cctctgtcatcacacgaatcgagatgctctgagatctgactacgcgcgacgaacctcgtctggc	480
Db	2502	ACCTGGCTGCTCGTGGAGACATAGAACATTCAGACGACCGCTCCGAGGGCTCACAGGCCGGG	2563
Qy	481	tgaccttaccgctcctactcgtgtgcatc-----aacgctgtcagacaccacgacacatcgg	534


```
Db 2562 GGTCCATCCGTCGCCAACGCGATGCGGCAATGCGGCAACGGCCACGACACGACG 2621
QY 535 tctatgccttgaagaaccctgtgtctcagcagtcgagcgagcgagcctcaacc 594
Db 2622 TGTGGGACACTCAACGCCCTGTTCCACGACGCCAATACGGCGCGCGTAAGCCCTGCAT 2681
QY 595 tgcctgcgcgcgtcagcagcgagcctatgacagctctcgcgagcgagatcactcagggc 654
Db 2682 ACATGCGCGGCGTGGCGGAGGCGCTTCGGCGCGCTGAGCGGCAACCGCATTTACTCAG 2741
QY 655 accgagtcacaaagaaccgcgtcagacgcgcacccatcgaccctctggggcccgagag 714
Db 2742 GCGTCATGACAAAACCCCGGACATCGCTGGGAAAGGAATGGCTCCACTCAATC 2801
QY 715 cgcctcagagcgtcgccctcgcaacacccctcgacagatcaccagcactcgagag 774
Db 2802 TCTACACACTCAGCCACATGAGGCGGAGCTCGGCGGAAACATGCCACCCCGCGCTGGC 2861
QY 775 cagggaaccgcgtcgcaacgctacccgatacaagctcgagcgagcgaacgtcactgtcg 834
Db 2862 GTGAGGACAGCACGTAACAAAGCGGCTCCGACGCGCTAGGGCGGAATTGCGCACTGTTCG 2921
QY 835 acacacccgcagatgtggcctaccgagcgctgc 867
Db 2922 ATTCCGTCAGGTTGTGGGCTATGTCGCCGCC 2954
```

RESULT 13
US-08-316-950-12
Sequence 12, Application US/08316950
Patent No. 5679515

GENERAL INFORMATION:
APPLICANT: STOVER, Charles K.
TITLE OF INVENTION: HICKEY, Mark J.
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316, 950
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 15371A-001100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 6047 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (plasmid)

FEATURE:
NAME/KEY: misc.feature

LOCATION: 1..6047
OTHER INFORMATION: /standard_name= "plasmid PMH30-lux"
US-08-316-950-12

```
Query Match 2.3%; Score 80.6; DB 1; Length 6047;  
Best Local Similarity 47.6%; Pred. No. 5e-10;  
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;  
QY 301 agagctgtcgtcagcagcagcgtcgtcagcagcagcagcagcagcagcagcagcagc 360  
Db 2405 AGCTTGCTGCTGCGCTAGTGGCGCTGGCAAGCAGCATCTGCTCAGAGGAGATCTACGCC 2464  
QY 361 tgactgcgcagagagcgctgagatgcttaccatcgaaggaaccgctgtgatgagt 420  
Db 2465 AAAGCGCGGCGTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCG 2524  
QY 421 ccttgatcaccagatcagatgcttgcagatgcttgcagatgcttgcagatgcttgcag 480  
Db 2525 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2584  
QY 481 tgcctcaccgtcctcagctgctcagctgctcagctgctcagctgctcagctgctcag 534  
Db 2585 GGTCCCATCCGTCGCCAACGCGATCGTGGGCAATCGCGCAACGCGCGCACACGACGAG 2644  
QY 535 tctatgccttgaagaaccctgtgtctcagcagtcgagcgagcgagcctataacc 594  
Db 2645 TGTGGGACACTCAACGCCCTGTTCCACGACCGAATGCGCGCGCTGTAACCGCTGCAT 2704  
QY 595 tgcctgcgcgcgtcagcagcgagcctatgacagctctcgcgagcgagatcctcagggc 654  
Db 2705 ACATGGCGGCGTGGCGGAGGCTTGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCG 2764  
QY 655 accgagtcacaaagaaccgcgtcagacgcgcacccatcgaccctctggggcccgagag 714  
Db 2765 GCGTCATGACAAAACCCCGGACATCGCTGGGAAAGGAATGGCTCCACTCAATC 2824  
QY 715 cgcctcagagcgtcgccctcgcaacacccctcgacagatcaccagcactcgagag 774  
Db 2825 TCTACACACTCAGCCACATGAGGCGGAGCTCGGCGGAAATGCCACCCCGCGCTGGC 2884  
QY 775 cagggaaccgcgtcgcaacgctacccgatacaagctcgagcgagcgaacgtcactgtcg 834  
Db 2885 GTGAGGACAGCACGTAACAAAGCGGCTCCGACGCGCTAGGGCGGAATTGCGCACTGTTG 2944  
QY 835 acacacccgcagatgtggcctaccgagcgctgc 867  
Db 2945 ATTCCGTCAGGTTGTGGGCTATGTCGCCGCC 2977
```

RESULT 14
PCT-US95-12642-12
Sequence 12, Application PC/TUS9512642

GENERAL INFORMATION:
APPLICANT: Pathogenesis Corporation
TITLE OF INVENTION: Seattle, Washington 98119
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12642
FILING DATE: 02-OCT-1995
CLASSIFICATION:


```

Db 2742 GCCTCATGACAAAACCCCGGCCACATCGCCTGGGAAACGGAATGGCTCCACTCAGATC 2801
QY 715 cgcctacgagctgagcgccctcgacacacccctcgacgagatccaagcactgcccgaag 774
Db 2802 TCTACACACTCAGCCACATCGAGGCGGAGCTGGGCGGAAACATGCCACCGCGCGCTGGC 2861
QY 775 cagggaaccgagctcgcaacgtaaccgatacaaggtcgccgcaacgtaaccctgttcg 834
Db 2862 GTCAGCAGACCACTACAAAGCGGCTCCGACGCGCTAGGCGGGAATTGCGACTGTTCG 2921
QY 835 acaccaccgcatgtgggataaccgggcccgtcc 867
Db 2922 ATTCCGTCAAGTTGTGGGCTATCGTCCCGCC 2954

```

Search completed: October 6, 2002, 16:34:24
 Job time: 17542 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:07 ; Search time 55.08 Seconds

(without alignments)
611.027 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596
Sequence: 1 MDSFETLFPESWLPKRKPLAS.....IAGSASRAHTRQOFLGLS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

_A_Geneseq_032802:*
1: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1596	100.0	303	21	AAV44636
2	329	20.6	368	18	AAW14835
3	315	19.7	368	13	AAW20992
4	287.5	18.0	367	14	AAW35443
5	218	13.7	248	13	AAW29622
6	218	13.7	248	13	AAW37872
7	106.5	6.7	458	22	AAW35246
8	97.5	6.1	492	22	AAW35243
9	97	6.1	3232	22	AAW69873
10	96.5	6.0	492	22	AAW35245
11	95.5	6.0	492	22	AAW35244

12	89.5	5.6	610	22	AAW90741
13	89	5.6	354	22	AAW10136
14	89	5.6	838	22	AAW70548
15	88	5.5	905	22	AAW29396
16	87	5.5	350	22	AAW61038
17	87	5.5	706	22	AAW48553
18	86.5	5.4	392	22	AAW53468
19	86	5.4	489	22	AAW86250
20	85.5	5.4	473	22	AAW16462
21	85.5	5.4	1219	22	AAW12658
22	85.5	5.4	1279	22	AAW78969
23	85.5	5.4	1301	22	AAW79953
24	85.5	5.4	1907	22	AAW15511
25	85.5	5.4	1909	22	AAW15534
26	85.5	5.4	2013	22	AAW12657
27	84.5	5.3	1018	22	AAW63718
28	84.5	5.3	4292	22	AAW17060
29	84.5	5.3	4302	17	AAW00870
30	84.5	5.3	4302	19	AAW33396
31	84.5	5.3	4302	19	AAW23830
32	84.5	5.3	4302	21	AAW92719
33	84.5	5.3	4303	17	AAW90302
34	84.5	5.3	4303	21	AAW95558
35	84.5	5.3	4339	17	AAW87539
36	84	5.3	693	22	AAW25744
37	84	5.3	693	22	AAW28822
38	84	5.3	717	22	AAW71311
39	83.5	5.2	299	21	AAW91671
40	83.5	5.2	381	20	AAW60162
41	83.5	5.2	450	19	AAW80605
42	83.5	5.2	450	21	AAW94307
43	83.5	5.2	450	22	AAW37841
44	83.5	5.2	450	22	AAW01090
45	83.5	5.2	458	22	AAW25289

ALIGNMENTS

RESULT 1
AAV44636 standard. Protein: 303 AA.
AAV44636:
07-APR-2000 (first entry) *bovs SEA 2 have very function*
Proprionibacterium LMG 16545 protein-1.
DE Proprionibacterium LMG 16545; vector; plasmid; antigen; vaccine; enzyme;
KW Proprionibacterium LMG 16545; growth factor; clotting factor; antimicrobial; drug;
KW nutritional factor; vitamin B12; animal feed; lactic acid bacteria; foodstuff;
KW cheese.
OS Proprionibacterium freudenreichii LMG 16545.
WO9967356-A2.
29-DEC-1999.
PD 25-JUN-1999; 99WO-EP04416.
PF 25-JUN-1999; 98EP-0305033.
PR 25-JUN-1998; 98EP-0305033.
PA (KONN) GIST-BROCADES BV.
XX Pouwels PH, Van Luitjck N, Jore JPM, Luiten RGM;
XX WPI: 2000-136977/12.
XX N-PSDB; AA49691.
XX Novel vectors containing Proprionibacterium sequences, used to express
PT homologous or heterologous proteins


```

XX OS Synthetic.
XX PN MO9201783-A.
XX PD 06-FEB-1992.
XX PF 09-JUL-1991; 91WO-US04833.
XX PR 16-JUL-1990; 90US-0553907.
XX PA (YESH ) EINSTEIN A COLLEGE.
XX PI (UPI-) UNIV OF PITTSBURGH.
XX PI Jacobs WR, Hatfull G;
XX DR WPI: 1992-064943/08.
XX PT DNA site-specific integration into Mycobacteria - useful as
XX PT adjuvant in vaccines and as therapeutic agent for malaria,
XX PT influenza, herpes and human immunodeficiency virus
XX PS Example 3; Fig 24; 82pp; English.
XX CC Plasmid pYUB12 was constructed by ligating a 5kb Sau3 fragment from
XX CC pAL5000 (contg. a replication origin from M.fortuitum) to BamHI-
XX CC digested pJ666 (contg. an E.coli ori and kanamycin-neomycin
XX CC resistance sequences). A 2586bp HpaI-EcoRV fragment from pYUB12
XX CC (comprising the minimum sequence necessary for plasmid replication
XX CC in BCG) was ligated to PvuII-digested pYUB8 to form pYUB53. (Plasmid
XX CC pYUB8 is a pBR322 deriv. which includes an E.coli replicon and a
XX CC kanamycin resistance gene). Twelve restriction sites were removed
XX CC by digesting pYUB53 with AatI, EcoRV and PstI. To eliminate 792
XX CC bases of the (inactivated) tet gene, the plasmid was digested with
XX CC MarI and the 6407bp gel purified fragment was religated/
XX CC circularised. The resulting plasmid was designated pMV101.
XX CC The nucleotide sequence of plasmid pMV101 is printed in the
XX CC specification but the copy quality is too poor to allow the
XX CC sequence to be included on the GENESQ nucleotide database.
XX CC Three proteins are encoded by pMV101, i.e. the kanamycin resistance
XX CC protein and proteins "d" and "e" encoded by the complementary
XX CC strand. See also AAR20991 and AAR20993.
XX SQ Sequence 368 AA;

Query Match 19.7%; Score 315; DB 13; Length 368;
Best Local Similarity 33.2%; Pred. No. 2e-24; Mismatches 128; Indels 42; Gaps 12;
Matches 104; Conservative 39;

OY 2 DSFETLPESWLPKRRPLASAE-KSGAYRHVTRORALELPYIEANPLVMOSLYITDSD 60
DB 66 defeql-----wlpypplaeaddlleglyrq-srasalgrrryieanplalnnllvddndp 120
OY 61 ADWADLA--GLPSPSYVMNRVTGTHIVALKNPVCLTDARRRPNILARVEGLCD 118
DB 121 aalralsaagsbplpnaivgrangahavvalnapvpteyartrkplaymaaceaglr 180
OY 119 VLGDSASYGHRITKNPLSTAHNTLWGPADALVELRALHTLDEIHA-LPEAGNPR----- 172
DB 181 avdgdrsygmltkwpgelawetewlhdsd-----lytlshleaeelgampmp-----prwrgt 233
OY 173 -ANVRSYTVGRVNTLFDTRMAYRAVRHSMGCPVAEME---HTVEHILHNTERTIADE 228
DB 234 tynaaptrpgrncalfdsrlywayrpalmrlylptnvnvlgylralyaeacrranaeprnd 293
OY 229 FAYGPLGLNELKHLRSISRW-----VWRN---FTPTFRARQKALISRGASGKGEGG 279
DB 294 vcpgrlpdpdeavralanslwriltsrliwadgjlvyaealrsarqsalsrkgsa----- 346
OY 280 HKGSIASGASRRA 292
DB 347 -artaastlvarta 358

```

```

RESULT 4
AAR34543
ID AAR34543 standard; Protein; 367 AA.
XX AC AAR34543;
XX DT 04-JUN-1993 (first entry)
XX DE Tet d gene product from pMV101 - a deletion mutant of pYUB125.
XX KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
XX KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
XX KW pertussis; malaria; influenza virus; CTL; herpes virus.
XX OS Mycobacterium.
XX FH Key Location/Qualifiers
XX FT Misc-difference 269
XX FT /note= "X undefined in specification"
XX FT Misc-difference 317
XX FT /note= "X undefined in specification"
XX PN WO9307897-A.
XX PD 29-APR-1993.
XX PF 21-OCT-1992; 92WO-US09075.
XX PR 21-OCT-1991; 91US-0780261.
XX PA (MEDI-) MEDIMUNE INC.
XX PI Stover CK;
XX DR WPI: 1993-152187/18.
XX DR N-PSDB; AA041316.
XX PT Expression vector for expressing protein or polypeptide in
XX PT mycobacterium - contg DNA sequences encoding lipoprotein
XX PT secretion signal and peptide heterologous to bacteria expressing
XX PT fusion protein of lipoprotein heterologous to bacteria
XX PS Example 1; Fig 5; 86pp; English.
XX SQ Sequence 367 AA;

Query Match 18.0%; Score 287.5; DB 14; Length 367;
Best Local Similarity 33.3%; Pred. No. 1.6e-21;
Matches 98; Conservative 35; Mismatches 124; Indels 37; Gaps 11;

OY 2 DSFETLPESWLPKRRPLASAE-KSGAYRHVTRORALELPYIEANPLVMOSLYITDSD 60
DB 66 defeql-----wlpypplaeaddlleglyrq-srasalgrrryieanplalnnllvddndp 120
OY 61 ADWADLA--GLPSPSYVMNRVTGTHIVALKNPVCLTDARRRPNILARVEGLCD 118
DB 121 aalralsaagsbplpnaivgrangahavvalnapvpteyartrkplaymaaceaglr 180
OY 119 VLGDSASYGHRITKNPLSTAHNTLWGPADALVELRALHTLDEIHA-LPEAGNPR-----RRN 174
DB 181 v-dgdrsygmltkmpgihawetew-----lhedytlshleaeelgampmprrwrgt 231

```


XX Corynebacterium thermoaminogenes.
XX
XX EPI076094-A2.
XX
XX PD 14-FEB-2001.
XX
XX PF 11-AUG-2000; 2000EP-0117225.
XX
XX PR 12-AUG-1999; 99JP-0228391.
XX
XX PA (AJIN) AJINOMOTO CO INC.
XX
XX PI Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;
XX Sugimoto S;
XX
XX DR WPI: 2001-193202/20.
XX
XX DR N-PSDB: AAF24442.
XX
XX PT New plasmids derived from Corynebacterium thermoaminogenes, useful for
XX improving coryneform bacteria, which can grow at elevated temperatures,
XX and for producing useful substances (e.g. L-amino acids)
XX
XX PS Claim 5; Page 20-21; 32pp; English.
XX
XX CC The present invention provides plasmids isolated from Corynebacterium
XX thermoaminogenes encoding a rep protein. These are useful for improving
XX Coryneform bacteria, which are capable of growing at elevated
XX temperatures and of producing substances such as L-amino acids. The
XX present sequence is the rep protein from a Coryneform bacterial plasmid.
XX
XX SQ Sequence 458 AA;

Query Match 6.7%; Score 106.5; DB 22; Length 458;
Best Local Similarity 20.1%; Pred. No. 0.026;
Matches 60; Conservative 42; Mismatches 119; Indels 77; Gaps 14;
QY 72 SPSYVSMNRVTTGGHIVYALKNPVCLTDAARRRPNLRLARVEQGLCDVYLGDPASYGHRIT 131
DB 137 apawlgvnpplsgkqqlwci-dprvfaegtsntrlllaatteentrfmgadqatshls 195
QY 132 KNPLESTAH---ATLM---GPADALYELRALAHTL-----DEIHALPEAGNPRRNWTR 177
DB 196 lwpvlhvsddplayswhcqhnrvdldelmevartmtgskkprehahgfsgrarleaar 255
QY 178 STV--GRNVTLFDTTRMAYRA-----VRHSWG--GPVAMENITVEHT----- 217
DB 256 kataeakaalaadatlptaleasgqlldgvrvlwaagraar-detaifhalvtvgylka 314
QY 218 --HLNETIIDEF-----ATGPLG-----LNETKHLRSISRWVWNPETFR 260
DB 315 agerlkdkkildayeraynvagavagdgrepdlpamrdtqtnarrrayvakg----- 367
QY 261 ARQKAISLRGSKGKEG-----GKGGIASGASRRRA-----HTROQFLGLS 303
DB 368 --qptvsarsltqtsgrgrkalatmgrrgaatsnarrrwadpsdyarqrerlarams 423

RESULT 8
AAB35243
ID AAB35243 standard; Protein: 492 AA.
XX
XX AC AAB35243;
XX
XX DT 08-MAY-2001 (first entry)
XX
XX DE Corynebacterium thermoaminogenes plasmid pYU1 rep protein.
XX
XX KW Plasmid: pYU1; rep; coryneform bacterium; L-amino acid; thermostable;
XX high temperature.
XX
XX OS Corynebacterium thermoaminogenes.

XX
XX EPI076094-A2.
XX
XX PD 14-FEB-2001.
XX
XX PF 11-AUG-2000; 2000EP-0117225.
XX
XX PR 12-AUG-1999; 99JP-0228391.
XX
XX PA (AJIN) AJINOMOTO CO INC.
XX
XX PI Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;
XX Sugimoto S;
XX
XX DR WPI: 2001-193202/20.
XX
XX DR N-PSDB: AAF24439.
XX
XX PT New plasmids derived from Corynebacterium thermoaminogenes, useful for
XX improving coryneform bacteria, which can grow at elevated temperatures,
XX and for producing useful substances (e.g. L-amino acids)
XX
XX PS Claim 1; Page 11-12; 32pp; English.
XX
XX CC The present invention provides plasmids isolated from Corynebacterium
XX thermoaminogenes encoding a rep protein. These are useful for improving
XX Coryneform bacteria, which are capable of growing at elevated
XX temperatures and of producing substances such as L-amino acids. The
XX present sequence is the rep protein from a Coryneform bacterial plasmid.
XX
XX SQ Sequence 492 AA;

Query Match 6.1%; Score 97.5; DB 22; Length 492;
Best Local Similarity 20.2%; Pred. No. 0.26;
Matches 55; Conservative 42; Mismatches 102; Indels 73; Gaps 14;
QY 72 SPSYVSMNRVTTGGHIVYALKNPVCLTDAARRRPNLRLARVEQGLCDVYLGDPASYGHRIT 131
DB 137 apawlgvnpplsgkqqlwci-dprvfaegtsntrlllaatteentrfmgadqatshls 195
QY 132 KNPLESTAH---ATLM---GPADALYELRALAHTL-----DEIHALPEAGNPRRNWTR 177
DB 196 lwpvlhvsddplayswhcqhnrvdldelmevartmtgskkprehahgfsgrarleaar 255
QY 178 STV--GRNVTLFDTTRMAYRA-----VRHSWG--GPVAMENITVEHT----- 217
DB 256 kataeakaalaadatlptaleasgqlldgvrvlwaagraar-detaifhalvtvgylka 314
QY 218 --HLNETIIDEF-----ATGPLG-----LNETKHLRSISRWVWNPETFR 260
DB 315 agerlkdkkildayeraynvagavagdgrepdlpamrdtqtnarrrayvakg----- 367
QY 261 ARQKAISLRG-----ASKG-----GKGGHK 261
DB 368 --qptvsarsltqtsgrgrkalatmgrrgrgqk 397

RESULT 9
ABB69873
ID ABB69873 standard; Protein: 3232 AA.
XX
XX AC ABB69873;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 36411.
XX
XX KW Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.

[illegible]

XX	14-FEB-2001.
PD	
XX	11-AUG-2000; 2000EP-0117225.
XX	
XX	12-AUG-1999; 99JP-0228391.
PR	
XX	(AJIN) AJINOMOTO CO INC.
PA	
XX	Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;
PI	Sugimoto S;
PI	
XX	WPI; 2001-193202/20.
DR	
DR	N-PSDB; AAF24441.
XX	
XX	New plasmids derived from Corynebacterium thermoaminogenes, useful for
PT	improving coryneform bacteria, which can grow at elevated temperatures,
PT	and for producing useful substances (e.g. L-amino acids) -
XX	
PS	Claim 4; Page 17-18; 32pp; English.
XX	
CC	The present invention provides plasmids isolated from Corynebacterium
CC	thermoaminogenes encoding a rep protein. These are useful for improving
CC	Coryneform bacteria, which are capable of growing at elevated
CC	temperatures and of producing substances such as L-amino acids. The
CC	present sequence is the rep protein from a Coryneform bacterial plasmid.
XX	
Sequence	492 AA;

Query Match	6.08	Score	96.5	DB	22	Length	492
Best Local Similarity	21.08	Pred. No.	0.33				
Matches	57	Conservative	41	Mismatches	101	Indels	73
						Gaps	15

	Query Match	6.1%;	Score 97;	DB 22;	Length 3232;
	Best Local Similarity	22.3%;	Pred. No. 4.7;		
	Matches	49;	Conservative	35;	Mismatches 70; Indels 66; Gaps 11.
QY	101 ARRPIN--LIARVHQL-----CDVIGDASTGHRITKNP---LSTAAATLWG	144			
	: : : : : : : : : : : : :				
Db	3009 akrraagallfegbqgsrlregldvrcsgslggsdahnghlessapkpwlssagsglsg	3068			
QY	145 PA-----DALYEL---RALAHNLDELHIA-LPEAGNRNRVVRTRGVARNLTLPFT	189			
	: : : : : : : : : : : :				
Db	3069 dstkvttrsvvpdaiaeycdsdeashqfgygarstclpslmpnq-vaast-ns pr pt	3125			
QY	130 TRMAYRAVRHWSGSPVAEMEHVTEHILNLNETIADFEATPGLNELKHLSHSISHW	249			
	: : : : : : : : : : : :				
Db	3126 tgv-----gprlctprtkqg-----salnhtpnlst-	3153			
QY	250 VWRNTPETFRAROKAISLRGASKSGKEGHHKGIGASGS	289			
	: : : : : : : : : : : :				
Db	3154 vsatgsaksartryspglaraaavsgaagsggyvavgs	3193			

Qy	72	SPSYVSMNRVTTTGGIIVVIALKNPVCILTAAARRPINLILARVQGLCDVYGGCASVGHRTT	131
Db	137	apawagvnpplagkqclvnci-dpvfaaegtsntcrlllaatteenmtrivgaqafshrs	195
Qy	132	KNPLUSTAH---ATLM---GPADALYELRALAHTL-----DEIHALPE--AGNPRRNVR	177
Db	196	lwpilhsddptcayshcqhnrvtldelimevartmtgskprkhhqefssgrarleaar	255
Qy	178	STV--GRVVTFEDTTRMMAYRA-----VHRSWG--GPVAEMHTYPEHI-----	217
Db	256	kataeakaalaadatlpraleasgdlldgvrvlwaegaaar-detafnhalvtvyqlka	314
Qy	218	-HLLNETTIADDF-----ATGPLG-----LNELKHLRSRISRWVWNNPPEPFR	260
Db	315	agerlkdaakldayeraaynvagavgadgpreplpmrdrctlmarrvrayvagy-----	367
Qy	261	ARQKALISLRG-----ASKG-----GKGGGKH 261	
Db	368	-qpvtvaarsltqtsrgrkalatlmngrryqk 397	

	RESULT	10
AAB35245		
ID	AAB35245	standard; Protein: 492 AA.
XX		
AC	AAB35245;	
XX		
DT	08-MAY-2001	(first entry)
XX		
DE	Corynebacterium thermoaminogenes plasmid pYM3 rep protein.	
XX		
KW	plasmid; pYM3; rep; coryneform bacterium; L-amino acid; thermostable;	
KW	high temperature.	
XX		
OS	Corynebacterium thermoaminogenes.	
XX		
RN	EPI076094-A2	

XX	AB35244	standard; Protein: 492 AA.
XX	AB35244	
AC	AB35244	
XX		
DT	08-MAY-2001	(first entry)
XX		
DE	Corynebacterium thermoaminogenes	plasmid pYW2 rep protein.
XX		
KW	plasmid; pYW2; rep; coryneform bacterium; L-amino acid; thermostable;	
KW	high temperature.	
XX		
OS	Corynebacterium thermoaminogenes.	
XX		
PN	EP1076094-A2.	
XX		
PD	14-FEB-2001.	

```
XX 11-AUG-2000; 2000EP-0117225.
PF
XX
XX 12-AUG-1999; 99JP-0228391.
PR
XX
XX (AJIN ) AJINOMOTO CO INC.
PA
XX
XX Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;
PI Sugimoto S;
XX
DR WPI: 2001-193202/20.
DR N-PSDB; AA24440.
XX
XX
XX New plasmids derived from Corynebacterium thermoaminogenes, useful for
PT improving coryneform bacteria, which can grow at elevated temperatures,
PT and for producing useful substances (e.g. L-amino acids) -
XX
XX
XX Claim 4; Page 14-15; 32pp; English.
PS
XX
XX The present invention provides plasmids isolated from Corynebacterium
CC thermophilum encoding a rep protein. These are useful for improving
CC coryneform bacteria, which are capable of growing at elevated
CC temperatures and of producing substances such as L-amino acids. The
CC present sequence is the rep protein from a Coryneform bacterial plasmid.
XX
XX
XX Sequence 492 AA;
SQ
XX
XX
XX Query Match 6.0%; Score 95.5; DB 22; Length 492;
Best Local Similarity 21.2%; Pred. No. 0.42;
Matches 58; Conservative 42; Mismatches 97; Indels 77; Gaps 16;
XX
XX 72 SPSVSMNVTGGHIVYALKNPVCLTDARRRPNILARVEGGLCDVLSYGRIT 131
DB 137 apawgvnplsgcqlwcl-dpvaegtsnttrlaatteemrrvgadafahris 195
QY 132 KNP-----STAHATW---GPADALVELRALAHTL---DEIHALPE--AGNPRNV 175
DB 196 rplihvfdpdtays--whcqhnrvdlldelmevarmtgskprkhaqefssgrarlea 253
QY 176 TSTV--GNVTLFDTRMAYRA-----VHSMG--GPVEMHTVFEHT----- 217
DB 254 arkataeaalaladclptaleasgdlldgyrvlwaegraar-detafrhalvgyql 312
QY 218 ----HLNFTIADFE-----ATGPLG-----LNELKHSRSISRWRNFTPER 258
DB 313 kaagerlkaklidayerayvavagvagdgrepdlpamdrqumarrvayvakg----- 367
QY 259 FRARQKAISLRG---ASKG-----GREGHK 281
DB 368 ----qptvaarsltqtsrgkcalatmrggqk 397
XX
XX
XX RESULT 12
AAG90741
ID AAG90741 standard; Protein; 610 AA.
XX
XX AAG90741;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum protein fragment SEQ ID NO: 4495.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
```

```
XX 16-DEC-1999; 99JP-0377484.
PR
XX 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOWA ) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX
XX WPI: 2001-376931/40.
DR N-PSDB; AAH65960.
XX
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Claim 17; SEQ ID NO: 4495; 246pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
XX Sequence 610 AA;
SQ
XX
XX
XX Query Match 5.6%; Score 89.5; DB 22; Length 610;
Best Local Similarity 21.2%; Pred. No. 2.5;
Matches 68; Conservative 37; Mismatches 115; Indels 101; Gaps 16;
XX
XX 24 SGAYRHVYTORALEPLYEANPLVMQS---LVITD--RDASDQWADLAGLSPSYVSM 78
DB 48 sytdengt-----pilygadkegyvqladkynqiveditgl-glydli 93
QY 79 NRTVTGHIY-----YALKNPV--CLTDARRRPNILARVEGGLCDVLSG 123
DB 94 tttlnhyavvqelfrgjldngymikettlgaistsgr--clpdrylegtcpicgtd 150
QY 124 ASYGHRTKNPLSTAHATWGPADALY---ELRALAHTLDEIHALPEAGNPRNVRTST 179
DB 151 gargdgcncngqldpadlinpyskingetpeltetehfidlpalaea-----lteml 204
QY 180 VGR-----NVTLFDTTRMAYRAVRH--SWG--GPVEMHTVFEHTHLNFTII 225
DB 205 kgredwrvpnlkfslhlditp---tamsrdldwglpivpegvqdnakklyvfdavv 261
QY 226 ADEFATGPLGLNELKHSISRWRNFTPE-----TFRARQ 263
DB 262 g-----ylsasl-ewayrsgdpeawrtfmdpetksyufmgkndithsqi 306
QY 264 KAISLRGASKGREGHKGI 284
DB 307 wpaellgyagkgarggeigldl 327
XX
XX
XX RESULT 13
AAE10136
ID AAE10136 standard; Protein; 354 AA.
XX
XX AAE10136;
AC
XX
XX 29-NOV-2001 (first entry)
DT
```

xx		Streptomycetes noursetl nystatin gene, ORF2.
DE		
xx		Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM		antifungal; antibiotic; transcriplional activator.
RW		
xx		
OS		Streptomycetes noursetl.
xx		
PT		MO200159126-A2.
PD		
xx		16-Aug-2001.
PF		
PE		08-FEB-2001; 2001WO-GB00509.
PR		
PPR		08-FEB-2000; 2000GB-0002840.
PR		10-Apr-2000; 2000GB-0008785.
XX		14-Apr-2000; 2000GB-0009387.
PA	(UNO-)	UNITV NORGES TEKNISK NATURVITJENSKAPELIG.
PA	(SMTF)	SINTERE STEIJELEN IND TEK FORSK.
PA	(ALPH-)	ALPHAAMA AS.
PA	(SINV-)	SINVENT AS.
PA	(DZIEC)	DZIEGLEWSKA H.
PA	(ZOTEC)	ZOTCHEV S B.
PA	(SEKU)	SEKUROVA O N.
PA	(ETAE)	EJAERVIK E.
PA	(BRAU)	BRAUTASET T.
PA	(STRO)	STROM A R.
PI	Zotchev SB,	Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI	Valla S,	Ellingsen TE, Sletta H, Gulliksen O;
DR	WPL:	2001-557614/62.
XX	N-PSDB:	AAD17184.
PT	New nystatin polyketide synthase polynucleotides and polypeptides,	
PT	useful as antibiotics and antifungals -	
XX		
PS	Claim 15; Page 179; 266pp; English.	
CC	The present invention relates to the cloning and sequencing of the gene	
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme	
CC	involved in the biosynthesis of the macrolide antibiotic nystatin.	
CC	The nystatin PKS is useful as antifungal antibiotics. The present	
CC	sequence is a transcriptional activator encoding Streptomycetes noursetl	
CC	nystatin gene, ORF2.	
XX		
SO	Sequence	354 AA;
	Query Match	5.6%; Score 89; DB 22; Length 354;
	Best Local Similarity	25.5%; Pred. No. 1.3;
	Matches	79; Conservative 33; Mismatches 114; Indels 84; Gaps 16;
OY		
	16 KPLASAEKSGAVRNHYTRORALELPRIEANPLVMQSLVTIDRNASADMMA--DLAQLPSP 73	
	: : : : : :	
DG	93 ralaaprhvalshvtgdrldqlymactdpmalsrwcthdlaaldgvraatrlag-pvh 151	
OY		
	74 SYVSNNRYTTTG-HIVYAALKNPVCUTDAARR---PINLLARVEQGCDVLG--GDASYG 127	
	: : : : : :	
DG	152 tegsrwtirralgrtnvarl-----aaasrntcdprafvldeiqqltvalsvdgracy- 205	
OY		
	128 HRITKNPLSTAHATLGMGPADALYLELRALAHTLDDELTHALPEAGNPNRRNYTRSYGNVTLF 187	
	: : : : : :	
DG	206 -----talae----eqcgagpd--lvrrivqr-----lf 227	
OY		
	188 DTTKRMAYRAVRHSHGGGAWEKVTEFHILLNETITA--DEFATGPLGLNELNLHSRS 245	
	: : : : : :	
DG	228 aadmlha-----rcvearplsewptv---sfwgvpparilrevlrtygmrevlcasv 279	
OY		
	246 ISR-----HWNRNFETFRAR-----QKASISLGAKKG--KEGHKKGI 284	
	: : : : : :	
DG	280 isrlnhlavawvsrldaqgrtevrilaeraadltvetereavalwhmhkgghlldeeegyrvgv 339	

Oy	285	ASGASRRRAH	294
Db	340	tplalwrept	349
RESULT 14			
ID	ABB70548		
XX	ABB70548 standard; Protein:	838 AA.	
AC	ABB70548;		
Dt	26-MAR-2002	(first entry)	
DE	Drosophila melanogaster polypeptide SEQ ID NO 38436.		
KW	Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.		
OS	Drosophila melanogaster.		
PN	WO200171042-A2.		
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
PR	23-MAR-2000; 2000US-191637P.		
XX	11-JUL-2000; 2000US-0614150.		
PA	(PEKE) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	MP1: 2001-6556860/75.		
XX	N-FSDR; ABLI4651.		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -		
PS	Disclosure; SEQ ID NO 38436; 21pp + Sequence listing; English.		
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABLJ30511), expressed DNA sequences (ABLI01840-ABLI6175) and the encoded proteins		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence	838 AA;	
Query Match	5.6%;	Score 89;	DB 22; Length 838;
Best Local Similarity	22.8%;	Pred. No. 4.5;	Indels 118; Gaps 10;
Matches	64;	Conservative	29; Mismatches
OY	4 FETLEPESMLPRKPLASAEGAYRHVTRORALEPYE-----AN 44		
Db	377 ftlppr---lpqkptlssasssgvsevttyprhlevgsksgseevttvrypta 433		
OY	45 PLVMOSLVITDR-DASDDMAADLAGLPSPSTIVS-MNRVTGTGHIVALKNPCLTDAAR 102		
Db	434 psltsvdltssgssessvevfttpayfvgvtlleclsldytvpclpttpr 493		
OY	103 RRPINILARVGQCDVLGDGDASYHRTTKNPLSAHNATIMCPNALVE----- 151		
Db	494 vwpvrpfafapeppldvvetastahlwtevypttaapfftegyaeavlltthtsagrfit 553		

```

QY 152 -----REALNHTDENHLEPEANSPRNYT-----RSTVGRNLTLEDTTQMA 194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 vqpragvttsptrdswevrlprtpqivtvlidsmevpslitttgs-----ptchhh 608
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 YRAVYRHSNGS-----PVAEMETVEVENHNLNETLIIDEPAT 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 hhhprhhaeagtlqpleeden- hhhhh-----hdeft 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	15
ABG29396	
ID	ABG29396 standard; Protein; 905 AA.

PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

PS Claim 20; SEQ ID No 59755; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 905 AA;

Query Match	Similarity	5.5%	Score	88	DB	22	Length	905	
Best Local	Similarity	24.7%	Pred. No.	6.4					
Matches	63	Conservative	27	Mismatches	87	Indels	78	Gaps	15
QY	27	YRHVYRORALE-----LPVI-----EANNPIVMOS--LVITDRDSDADMAADLAG	69						
Db	171	fyimrrellinkmgtlpyvapellkrrfehnepadvscgilvltamlagelpw-----d	225						
QY	70	LPSPS---YVSNMRTTGTGHIVALKNPCLDAAARRRPIINLARQEGICVLGGDASY	126						
Db	226	qpsdscgeysdwkckett-----ylnpwkkidsa---plogscqne-----lsg--sl	268						
QY	127	GHRITKNPLSTAHATLWGPAD--ALYELRALAHITLDELHALPEAGNPRRNVRSTYGRNV	184						
Db	269	dhqtsvqqlpgeffslenspdsalayetsgintltise-hgssehgsskltvaehtsge--	325						
QY	165	TLFDTRRMATYAIVRHSWGCPV---AEWEHVFEEH-----IHLNETTII-----	225						
Db	336	-----haeshaesepatehaeghehtvgqpsgeqpsgehlsgeqblslelsge	375						
QY	226	--ADEFATGPIGLINE	238						
Db	376	qpsdeqpsgehngsge	390						

Search completed: October 4, 2002, 15:34:40
Job time: 93 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 28.21 Seconds

(without alignments)
1032.083 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596

Sequence: 1 MDSEFTLPESWLPKPLAS.....IAGASRRARHROQFLEGLS 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:**

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	329	20.6	336	2 JS0052	37K protein - Myco
2	282.5	17.7	297	2 S04455	hypothetical prote
3	282.5	17.7	304	2 S04456	hypothetical prote
4	198.5	12.4	248	2 S32701	hypothetical prote
5	100	6.3	99	2 S32700	hypothetical prote
6	98	6.1	1461	2 B70588	probable polyketid
7	97.5	6.1	1876	2 E97944	zinc metalloprotei
8	93.5	5.9	228	2 H72703	probable translati
9	89.5	5.6	606	2 T37217	probable ATP/GTP b
10	89	5.6	1216	2 T09224	spindle assembly c
11	89	5.6	4558	2 C82199	RTX toxin RtxA VCI
12	88.5	5.5	303	2 T36509	probable molybdopt
13	87.5	5.5	409	2 T14611	alanine transamina
14	87	5.5	457	2 S03321	regulatory protein
15	87	5.5	487	2 S60675	hypothetical prote
16	86.5	5.4	497	2 T14609	alanine transamina
17	86	5.4	672	2 JY4637	transketolase (EC
18	86	5.4	672	2 T03457	transketolase (EC
19	85.5	5.4	1217	2 T00270	hypothetical prote
20	85.5	5.4	1366	2 T35985	probable large pro
21	85	5.3	356	2 H98304	limb2-dependent al
22	85	5.3	356	2 AE2978	alkanesulfonate mo
23	85	5.3	409	2 T06767	cellulase (EC 3.2.
24	85	5.3	900	2 JH0157	mtag protein (Afl8
25	84.5	5.3	355	2 H98275	monooxygenase [imp
26	84.5	5.3	355	2 AE3008	probable integral
27	84.5	5.3	500	2 T36090	polycystic kidney
28	84.5	5.3	4302	2 A38971	proteinase II [imp
29	84	5.3	686	2 C90948	

30	84	5.3	686	2 G85796	proteinase II [imp
31	84	5.3	1693	2 AC3240	helicase, SNF2 fam
32	83.5	5.2	450	2 E95257	replicative DNA he
33	83.5	5.2	450	2 E98122	replicative DNA he
34	83.5	5.2	863	1 GNMHYH	retrovirus-related
35	83.5	5.2	1815	2 S73021	polyketide synthas
36	83.5	5.2	1822	2 F87203	polyketide synthas
37	83	5.2	251	2 I57999	SP-10 - western ba
38	83	5.2	285	2 I77964	probable acetolact
39	83	5.2	559	2 G83024	exodeoxyribonuclea
40	83	5.2	639	2 D82809	oligopeptidase B (
41	83	5.2	686	1 E64946	retrovirus-related
42	82.5	5.2	814	1 GNM5IP	cobn protein homol
43	82.5	5.2	1116	2 D97695	cobalamin biosynth
44	82.5	5.2	1116	2 AC2921	hypothetical prote
45	82	5.1	497	2 T47715	

ALIGNMENTS

RESULT 1
JS0052
37K protein - Mycobacterium fortuitum plasmid PAL5000
C:Species: Mycobacterium fortuitum
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 10-Dec-1999
C:Accession: JS0052; J01440
R:Raunzier, J.; Moniz-Pereira, J.; Gloquel-Sanzey, B.
Gene 71, 315-321, 1988
A:Title: Complete nucleotide sequence of PAL5000, a plasmid from Mycobacterium fortui
A:Reference number: JS0052; MID:89138007
A:Accession: JS0052
A:Molecule type: DNA
A:Residues: 1-336 <RNU>
A:Note: the authors translated the initiation codon GTG for residue 1 as Val
R:Labidi, A.; Mardis, E.; Roe, B.A.; Wallace Jr., R.J.
Plasmid 27, 130-140, 1992
A:Title: Cloning and DNA sequence of the Mycobacterium fortuitum var fortuitum plasm
A:Reference number: J01440; MID:92311251
A:Accession: J01440
A:Molecule type: DNA
A:Residues: 160-336 <LAB>
A:Cross-references: GB:M60875; NID:g149984; PIDN:AAA25372.1; PID:g149985
A:Experimental source: var. fortuitum
C:Genetics:
A:Genome: plasmid
A:Start codon: GTG
C:Function:
A:Description: this protein is involved in the initiation of plasmid replication
C:Superfamily: Mycobacterium hypothetical 37k protein

Query Match 20.6%; Score 329; DB 2; Length 336;

Best local similarity 33.5%; Pred. No. 5e-20;

Matches 104; Conservative 39; Mismatches 131; Indels 36; Gaps 10;

QY	2	DSEFTLPESWLPKPLASAE	KS	GA	RY	HT	RO	RA	LE	PT	EA	NP	LM	OS	LV	TD	R	ASD	60																													
DB	34	DEFEQ----	W	LP	W	PL	ASD	DL	E	GI	Y	Q	----	S	R	SA	L	GR	RY	EA	NP	LM	LV	VD	Y	D	H	P	88																			
QY	61	ADNAADLA	---	G	L	P	S	P	S	V	S	N	R	V	T	T	G	I	Y	A	L	K	N	P	C	L	D	A	A	R	R	P	I	N	L	A	R	E	G	I	C	D	118					
DB	89	AALRALASRGSH	PL	PA	NI	V	G	N	R	A	N	G	H	A	Y	A	L	N	A	P	P	R	E	Y	A	R	R	K	L	A	Y	M	A	C	A	G	L	R	148									
QY	119	VLCGDSYGRIT	K	N	P	L	S	A	H	A	T	L	M	G	P	A	L	Y	E	L	R	A	L	A	H	T	L	D	E	I	H	A	L	P	E	A	G	N	P	R	R	N	T	---	176			
DB	149	AVYDGRISYGL	M	T	K	N	P	G	H	I	M	E	T	E	M	L	S	D	---	L	T	S	H	I	E	A	E	L	G	A	N	M	P	P	R	R	O	Q	T	Y	K	204						
QY	177	---	K	S	T	V	R	N	T	L	D	T	T	G	M	A	Y	R	A	V	R	H	S	G	P	V	A	E	M	---	H	T	V	E	H	I	L	M	E	T	I	A	D	E	F	A	T	231
DB	205	AAPPLGRCAL	P	D	S	V	L	M	A	Y	R	P	A	L	M	R	I	L	P	T	R	N	V	D	L	G	R	A	I	Y	A	E	C	H	A	R	N	A	E	E	P	C	N	D	V	C	P	264

QY 232 GRLGELNELKHLRSISRW-----VMRN---FTPEFRAROKAISLRGASCKGCKGCKG 282
||| : : : : : ||| : : : : : ||| : : : : :
Db 265 GLPDSEVRAINSIMRWITTKSRIMADGIYYEATLSRKOSAIRKGAA-----AR 316
QY 283 GIASGASRRA 292
|| : |||
Db 317 TAASIVARRA 326

RESULT 2

S04455

hypothetical protein - Escherichia coli plasmid ColE2

C:Species: Escherichia coli

C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 11-Jan-2000

C:Accession: S04455

R:Yasuda, H.; Horii, T.; Itoh, T.

Mol. Gen. Genet. 215, 209-216, 1989

A:Title: Structural and functional organization of ColE2 and ColE3 replicons.

A:Reference number: S04455; MUID:89218922

A:Accession: S04455

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-297 <YAS>

C:Genetics:

C:Superfamily: Mycobacterium hypothetical 37K protein

Query Match 17.7%; Score 282.5; DB 2; Length 297;

Best Local Similarity 29.5%; Pred. No. 3.4e-16;

Matches 87; Conservative 49; Mismatches 130; Indels 29; Gaps 11;

QY 13 LPRKPLASAEKSGAARHYTRORALELPIYEAN-PLVMOSLYI-TDRDASDADMAADLAGL 70
||| : : : : : ||| : : : : : ||| : : : : :
Db 12 LPHKPYCTNDFAYGVRIILPKNIAIILARFIQONQPHALYWLPEVDVTRTASIDMSD--RNC 69

QY 71 PSPSYVMNRVTTTGHIVYALKNPVCLDAAARRPINLARVEGGLCDVLDGASYGHR 130
||| : : : : : ||| : : : : : ||| : : : : :
Db 70 PAPNTTVKNPRNGHMLLYALALPVRTAPDASASALRYAAALERLCKLGADVYSGLI 129

QY 131 TKNPLSTAHATLWGPADALYELRALAHTLDEI-HALPEAGNPRNRVTS-TVGRNVTLFD 188
||| : : : : : ||| : : : : : ||| : : : : :
Db 130 CKNP---CH-----PEMOEVEWREPEPYTLDELADYLDLSASARRSVKNGYGLGRNYHLFE 181

QY 189 TTRMAYAVARHWSGVPV-AEWEHTVFEHIIHLNETIADDEFATGRLGELNELKHLRSIS 247
||| : : : : : ||| : : : : : ||| : : : : :
Db 182 KYRKMAVRAIRQGW--PVESQWLDVAVIQREVMYNASLPV-----PLSPAECERAIGKSLA 233

QY 248 RWMVNRFTPEFRAROKAISLRGASCKGCKG-----GHKGSIASGASRRRAHTRO 236
||| : : : : : ||| : : : : : ||| : : : : :
Db 234 KYTHKRFSPESGSVAOAAARGRGTKSRRAVPTSARSLKPEWELGISRATYYRK 288

RESULT 3

S04456

hypothetical protein - Escherichia coli plasmid ColE3

C:Species: Escherichia coli

C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Jun-2000

C:Accession: S04456

R:Yasuda, H.; Horii, T.; Itoh, T.

Mol. Gen. Genet. 215, 209-216, 1989

A:Title: Structural and functional organization of ColE2 and ColE3 replicons.

A:Reference number: S04455; MUID:89218922

A:Accession: S04456

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-304 <YAS>

A:Cross-references: GB:D30055; NID:g487267; PIDN:BAA06293.1; PID:g808865

C:Genetics:

A:Genome: plasmid

C:Superfamily: Mycobacterium hypothetical 37K protein

Query Match 17.7%; Score 282.5; DB 2; Length 304;
Best Local Similarity 29.5%; Pred. No. 3.5e-16;
Matches 84; Conservative 47; Mismatches 127; Indels 27; Gaps 11;

QY 13 LPRKPLASAEKSGAARHYTRORALELPIYEAN-PLVMOSLYI-TDRDASDADMAADLAGL 70
||| : : : : : ||| : : : : : ||| : : : : :
Db 12 LPHKPYCTNDFAYGVRIILPKNIAIILARFIQONQPHALYWLPEVDVTRTASIDMSD--RNC 69

QY 71 PSPSYVMNRVTTTGHIVYALKNPVCLDAAARRPINLARVEGGLCDVLDGASYGHR 130
||| : : : : : ||| : : : : : ||| : : : : :
Db 70 PAPNTTVKNPRNGHMLLYALALPVRTAPDASASALRYAAALERLCKLGADVYSGLI 129

QY 131 TKNPLSTAHATLWGPADALYELRALAHTLDEI-HALPEAGNPRNRVTS-TVGRNVTLFD 188
||| : : : : : ||| : : : : : ||| : : : : :
Db 130 CKNP---CH-----PEMOEVEWREPEPYTLDELADYLDLSASARRSVKNGYGLGRNYHLFE 181

QY 189 TTRMAYAVARHWSGVPV-AEWEHTVFEHIIHLNETIADDEFATGRLGELNELKHLRSIS 248
||| : : : : : ||| : : : : : ||| : : : : :
Db 182 KYRKMAVRAIRQGWPA-RSQWLDVAVIQREVMYNASLPV-----PLSPPECEARAIGKSLA 234

QY 249 WVMNRFTPEFRAROKAISLRGASCKGCKGCKGSIASGA 288
||| : : : : : ||| : : : : : ||| : : : : :
Db 235 YTHKRFSPESGSVAOAAARGRGTKSRRAVPTSARSLKPEWELGISRATYYRK 278

QY 249 WVMNRFTPEFRAROKAISLRGASCKGCKGCKGSIASGA 288
||| : : : : : ||| : : : : : ||| : : : : :
Db 235 YTHKRFSPESGSVAOAAARGRGTKSRRAVPTSARSLKPEWELGISRATYYRK 278

RESULT 4

S32701

hypothetical protein 3 - Corynebacterium glutamicum

C:Species: Corynebacterium glutamicum

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995

C:Accession: S32701

R:Shi, S.L.; Wang, Z.X.; Deng, Y.; Zhu, H.; He, B.; Zheng, Z.X.

submitted to the EMBL Data Library, March 1993

A:Description: Complete nucleotide sequence of a plasmid pX210142 from Corynebacteriu

A:Reference number: S32699

A:Accession: S32701

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-248 <SHI>

A:Cross-references: EMBL:X72691

Query Match 12.4%; Score 198.5; DB 2; Length 248;
Best Local Similarity 27.4%; Pred. No. 3e-09;
Matches 68; Conservative 35; Mismatches 94; Indels 51; Gaps 10;

QY 73 PSYVMNRVTTT-----GHIVYALKNPVCLDAAARRPIN----- 107
||| : : : : : ||| : : : : : ||| : : : : :
Db 2 PSRISWSSTSRHSCVRCGTETAGGLTPWLKTPF---KRAKRRVGAARGAIYPRHVRQ 58

QY 108 -----LLARVEGGLCDVLDGASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEI 162
||| : : : : : ||| : : : : : ||| : : : : :
Db 59 AQAALAYAAAVTEGLRSDVDGKGYGLITKNPENTHAMSHV-VYDKLYTLDELRFWLEET 117

QY 163 HAL-PEAGNPRNRVTSRVGRNVTLFDTTMMAYAVARHWSGVPV-AEWEHTVFEHIIHLN 221
||| : : : : : ||| : : : : : ||| : : : : :
Db 118 GFMPESWKTTRKRSPIGLGRNCALEFESARWAVEIRHGHGDDPG-LGRSIQATQALN 176

QY 222 ETIADDEFATGRLGELNELKHLRSISRWV-----WRNFTPEFRAROKAISLRGASGG 275
||| : : : : : ||| : : : : : ||| : : : : :
Db 177 QELTSE-----PLPVAEVDQIARSIHRWITKSRMWD-GRVAVEATEFTTIO---ATRGK 227

QY 276 REGGHKG 283
||| : : : : : ||| : : : : : ||| : : : : :
Db 228 RSAEHRWG 235

QY 276 REGGHKG 283
||| : : : : : ||| : : : : : ||| : : : : :
Db 228 RSAEHRWG 235

RESULT 5

S32700

hypothetical protein 2 - Corynebacterium glutamicum

C:Species: Corynebacterium glutamicum

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Aug-1995

C:Accession: S32700; S32199


```

Db      91  -- HLMGSGVNAVYVLEPDSNAAGLAVSRSSNVG6 - ADAAULDGLCRKKLTLPPLAASAAALG 147
QY      144  GPADALYELRALAHTLDEIHALLPEAGNPRNRVNTSTVGSRVNTLEFDTTBMAYRAYRHSWG 203
Db      148  VEAVAVFESYK-----VSRAQVEGV---ETR--SYRV--EGWG 180
QY      204  GPVAEMHTVPEHILHNETIIDEAFATGPIGLNLT 240
Db      181  ----DVEIVPEPIPLRLVSTATELGGGRAGVEPLR 213

RESULT      9
T37217
Probable AMP/GMP binding protein - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37217
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21615
A:Accession: T37217
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-606 <OLI>
A:Cross-references: EMBL:AL031514; PIDN:CAA20595.1; GSPDB:GN00070; SCOEDB:SC2H4.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC2H4.01

```

```

Query Match 5.6%; Score 89.5; DB 2; Length 606;
Best Local Similarity 22.3%; Pred. No. 14;
Matches 67; Conservative 38; Mismatches 127; Indels 69; Gaps 16;

OY 30 VTROPALELPYEANPLVMQSLVITDRDASDADWADLAGLSPSPYSVMNRVTTGHTIV 89
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 ITQOOLLEEGEPANELVAQ----WDAASSTGYTKELSSYPALFT---LSAAGAVLD 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 90 ALKNVVCCLTDA-----RRRINILARVQGLDVIYGGASVGHRTKPLPETA 138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 EENNALVDLDSAGVATSGPVVDETSFTVSARVLDLSAKLAKRVGYEALVAGQASAG 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 139 HAT--LWG--PADALYE--LRALAHITDELIALPEAGNPRNRVTSYVGRNV--LFDTT 190
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 ESSNALVMKVPADGVYQWKFTTAVGADGKVQSAEA--PGGDIKFTDTWQVTFGVDAQ 516
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 191 RMWAVRAVRHSGSPV--AEWETVFEHILH--LINFIIINAEFACRPLDLNELKHSRIS 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 EPW-----EWTDPADSAKTE--TYGRILHYVGFDPDSEFASG----- 553
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 248 RMWENFETPEFTRAKOKAISLNGASKGKEGHNKG-----IASGASRRATHROQPLEG 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 -----FTTPQYAGALAVNG--RG--SRGGTGGHILFGALEELRVTMTGAMSADQVRSQVLGG 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 302 L 302
      :
Db 606 V 606

RESULT 10
T09224
Spindle assembly checkpoint protein SIDA - Emeritella nidulans
C:Species: Emeritella nidulans, Aspergillus nidulans
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09224
R:Efilmov, V.P.; Morris, N.R.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z16619
A:Accession: T09224
A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1216 <EF1>

```

A:Cross-references: EMBL:AF032987; NID:93136093; PID:93136096
A:Experimental source: strain R153
C:Genetics:
A:Gene: sldA
A:Introns: 331/3; 477/1
C:Keywords: microtubule; mitosis

Query Match	5.6%;	Score 89;	DB 2;	Length 1216;
Best Local Similarity	21.5%;	Pred. No. 37;		
Matches	66;	Conservative	42;	Mismatches 93;
				Indels 106;
				Gaps 15.

```

RESULT 11
C82199
RTX toxin RtxA VC1451 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82199
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gyi, M.L.; Dodson, R.;
charson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4558 <HEI>
A:Cross-references: GB:AE004223; GB:AE003852; NID:9655942; PIDN:AAF94608.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1451
A:Map position: 1

Query Match 5.6%; Score 89; DB 2; Length 4558;
Best Local Similarity 23.6%; Pred. No. 2e+02;
Matches 50; Conservative 20; Mismatches 76; Indels 66; Gaps 9;

OY 124 ASYGRITKNP-----ISTAHATIMGPADALY-----ELRALHTLDEIH 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2386 ASKGQPYKNPKRSSASANKHAEWMLTQHAEDQAWAITAPYTDNHNHTLSTKLASDKDELH 2445
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 164 ALPEAGNPRRVNTRSTVGR--NVTLFDFTBMAYRAVH-----SMGCPVAEWHNYEE 215
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2446 ALGETSLNENKQDQNVASITNTMLNDMLPYVALTERTNLLVQEGDDEFEVRAVPGTECD 2505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 216 HIHLNETIADDEFATGPIGLNELKHLRSRSRWWRNF-----TPEFRAROKAISLR 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 2506 -----STILED-----PEDAQKAIERFILANFDNEQMPDELFVNDKVIS-- 2550

QY 270 GASKGKGGHKGIGASGASRRATROOFLEG 301

Db 2551 -----HHEG-----RTHVLAQKVDG 2565

RESULT 12

T36509

probable molybdopterlin-guanine dinucleotide blosynthesis protein - Streptomyces coelicol

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36509

R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: 221608

A:Accession: T36509

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-303 <SAU>

A:Cross-references: EMBL:AL096822; PIDN:CAB46951.1; GSPDB:GN00070; SCQDB:SCGD3.29C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCQDB:SCGD3.29C

Query Match

Best Local Similarity 23.5%; Score 88.5; DB 2; Length 303;

Matches 75; Conservative 33; Mismatches 104; Indels 107; Gaps 18;

QY 2 DSEETLPESWLPKPKPLAS---AEKSGAYRHVTRORAL---ELPYEAMPV---MQSL 51

Db 44 DPRPTAPYWARDEPPGGGLALAGLRHTTAEVYVASADLPFL-AGPTVGRLLSAL 102

QY 52 VTIDRDA---SDADWADLAGLPSPSYSMNRVTTHGHYVALNPPVCLTPAARRPINTL 108

Db 103 AAGGADVLTLTDAD-----GRDP-----LVAAYRAP-----ALRRELA 137

QY 109 LARVEGLCDVLDGDAYGHRITKNPLSTAHATLMGPADALYELRALAHITDEIHALPE 168

Db 138 LARRKAG-----KGERADGAOLGLP-----LRLTGLAL-RLTRVPDA 174

QY 169 GNPARNVTRSTVGRNVTLFTTRMAYRAVHSGPYAEHEHTVFEEHILNETIAD 228

Db 175 -----VAAPFC-----DTW-DLATAARARIREGHVLDWMISA-- 206

QY 229 PANGPLCLN-----ELKHLRSISRWVWRNFTP-ETFRANQAKISLGASGKGEGCHK 281

Db 207 -AKDELGIDIDVDTGILLDLARDAHGVAPAPADLTFLVGYAA-----GRAGGP 256

QY 282 GGASGASRRATROOFLE 300

Db 257 EAVAEETAKKAALAAQNAE 275

RESULT 13

T14611

alanine transaminase homolog - Trypanosoma cruzi

C:Species: Trypanosoma cruzi

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Jun-2000

C:Accession: T14611

R:Andersson, B.; Aslund, L.; Pettersson, U.

submitted to the EMBL Data Library, March 1998

A:Description: 93.4 kb of complete sequence from chromosome 3 of Trypanosoma cruzi.

A:Reference number: 218159

A:Accession: T14611

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-409 <AND>

A:Cross-references: EMBL:AF052832; NID:g3063540; PID:g3063548; PIDN:AACT14084.1

C:Genetics:

A:Map position: 3

C:Superfamily: alanine transaminase

Query Match

Best Local Similarity 23.3%; Score 87.5; DB 2; Length 409;

Matches 80; Conservative 36; Mismatches 127; Indels 101; Gaps 16;

QY 15 RKPLASAE-----KSGAYRHVTRORALE-----LPYEAMPVMOQLVITDRD 57

Db 33 RAALATPEGKAKYPFSSIVYCNINPQALEGKPLTFEFRQVMSLDA-PFLEENKVSQY 91

QY 58 ASDADMAA-----DLACLPSPSYSMNRVTTGHYVAL-KNPVCLTDA 100

Db 92 PADVAVARREYLRIHIGRTGAYTDSAGYAFARDIVARQINRDEHIEKPLVASSIFLTDG 151

QY 101 ARRAPINLAVEGGLCDVLDGDAYGHR--ITKNPLSTAHATLMGPADALYEL----- 152

Db 152 A-SSGVRLILO-----VLVGDSADAVMPPIPOPLTAQTLTGLGTPAMYVLCEDKNW 203

QY 153 -RALAHITDEI---HALPEA-----GNPRNVTRSTVGRNVTLFTTRMAYRAV 198

Db 204 ALNVEELASVYDECVAKNNATPRVLVINPGNPGVLDROVMEAVAFCCDR----- 256

QY 199 RNSGCPYAEHEHTVFEEHILNETIADDEPATGELGELKHLRSISRWVWRNFTPET 258

Db 257 -----GIVLMADEVYQENVYAAGRFLS--FREYVLGL-----PAP 290

QY 259 FRAROKAISLGASGK--GKEGKHKGIGASGASRRATROOFLE 300

Db 291 YNTDTIVLASLSTSKGIIG-ECGRRGYFCLTNFPAPVREYVK 333

RESULT 14

S03321

regulatory protein ntrC - Rhodobacter capsulatus

N:Alternate names: regulatory protein nifH1

C:Species: Rhodobacter capsulatus

C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 02-Feb-2001

C:Accession: S03321; S37299; S34983

R:Jones, R.; Haselkorn, R.

MOL Gen. Genet. 215, 507-516, 1989

A:Title: The DNA sequence of the Rhodobacter capsulatus ntrA, ntrB and ntrC gene anal

A:Reference number: S03318; MUID:89218961

A:Accession: S03321

A:Molecule type: DNA

A:Residues: 1-457 <JON>

A:Cross-references: EMBL:X12359; NID:g46077; PIDN:CAA30922.1; PID:g46079

R:Kranz, R.

submitted to the EMBL Data Library, March 1993

A:Reference number: S37299

A:Accession: S37299

A:Molecule type: DNA

A:Residues: 1-457 <KRA>

A:Cross-references: EMBL:X72382; NID:g313162; PIDN:CA51075.1; PID:g313166

R:Poster-Hartnett, D.; Cullen, P.J.; Gabbert, K.K.; Kranz, R.G.

MOL. Microbiol. 8, 903-914, 1993

A:Title: Sequence, genetic, and lacZ fusion analyses of a nifH3-ntrB-ntrC operon in R

A:Reference number: S34980; MUID:93360820

A:Accession: S34983

A:Molecule type: DNA

A:Residues: 334-449 <FOS>

A:Cross-references: EMBL:X72382

C:Genetics:

A:Gene: ntrC; nifH1

C:Superfamily: nitrogen assimilation regulatory protein ntrC; response regulator homo

C:Keywords: DNA binding; nucleotide binding; P-loop; phosphoprotein; transcription re

F:5-114/Domain: response regulator homology <RRA>

F:143-336/Domain: RNA polymerase sigma factor interaction domain homology <SFI>

F:171-178/Region: nucleotide-binding motif A (P-loop)

F:53/Binding site: phosphate (asp) (covalent) #status predicted

Query Match

5.5%; Score 87; DB 2; Length 457;

Best Local Similarity 24.6%; Pred. No. 15;
Matches 75; Conservative 28; Mismatches 100; Indels 102; Gaps 16;

Search completed: October 4, 2002, 15:35:15
Job time: 127 sec

OY 5 ETLFESWLPKPLA--SAKSGAYRHVROALELPIY-----EANPLVMQSLVITDR 56
DB 132 ETSRPEG--TDLPVIGRTAMQALRYVARVMNADLPVIMGESGKSLIAKAI----H 185
OY 57 DASD-----ADMAADLAGLPSYSVMNV-----TTTGHTIV---ALK 92
DB 186 DFSDRRTLPFVVAQAADLLGADGPSLLARQGRLVFEDEVDYDDEFGRIYRMIDLAP 245
OY 93 NFV-----CLTDAARRRPINLARVEQGLCDVLGSDASYGHRITKNPLSTAHAT 141
DB 246 DPPRIMATTQVDLGLAMEGR-----RFRDLYYRLGG-----VT 280
OY 142 LMGPADALYELRALAHTLDEIHALPE--AGNPRNVTRSTVGRNVTLPDTRMMAYRAVR 199
DB 281 LAVP-----ALRRVEDIDPLAEHFLGRAERD-----GLGMRASFSAEMGLVRA-- 324
OY 200 HSMGCPVAEMETHVEHHLNET-----IIADEPATGPLGL-NELKHLSSISRW 249
DB 325 YAMPGNVROLENTVRLVVTASEEITRTVEFVLGNCPAVEPLGAGGEGKLSASIAH 384
OY 250 VWRNF 254
DB 385 LRRYF 389

RESULT 15

S60675
hypothetical protein C - Corynebacterium glutamicum plasmid pGAI
C:Species: Corynebacterium glutamicum
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 22-Oct-1999
C:Accession: S60675
R:Neavera, J.; Palek, M.; Hochmannova, J.; Abrahmova, Z.
submitted to the EMBL data library August 1995
A:Description: Complete nucleotide sequence of the cryptic plasmid pGAI from Corynebacte
A:Reference number: S60673
A:Accession: S60675
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-487 <NES>
A:Cross-references: EMBL:X90817; NID:g951006; PIDN:CA62330.1; PID:g951009
C:Genetics:
A:Genome: plasmid pGAI

Query Match 5.5%; Score 87; DB 2; Length 487;

Best Local Similarity 20.2%; Pred. No. 17; Mismatches 116; Indels 72; Gaps 13;
Matches 59; Conservative 45;

OY 72 SPSYVSMNRVTTHIYVYALKNPVCLTDAARRRPINLARVEQGLCDVLGSDASYGHRIT 131
DB 138 APAMIGVNPPLSGKQQLIW-LIDPVYAAAGKTSPPNMLLAATTEEMTRVFGADQAFSHRLS 196
OY 132 KNPULSTAH---ATLW---GPDALYELRALAHTLDEIHALPEAGNPRNV----- 175
DB 197 RMPLEHVSDDPTAYKWHCOHDRVDRDLADLEIART-----MTGSOKPKKYEIDDFSSGRA 250
OY 176 -----TRSTV-GRNVTLPDTRFMAYRA-----VRHSMGCPVAEMETHVEHT----- 217
DB 251 RLEAQRATAEKALAILDASLPSALDSGLIDGVRVLMWNPERRADETAFRHALTVGY 310
OY 218 -----HLINETIADDF-----ATGPLG-----LNEKHLSSISRWVNRNFTP 256
DB 311 QLKAAGERLKKAKIITDAVEAVNVNAQVAGADGERPDLAMRDLTMARVRGVYAKG-QP 369
OY 257 ETPRAROKAISIURGASGKGKGGKGTASGASRA-----HTROQFLE 300
DB 370 VVPARRVETOSRG-RKALATWGRGAATSNARRWADESKYAQETRORLAE 420

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 15.85 Seconds

(without alignments)
740.191 Million cell updates/sec

Title: US-09-720-583A-2

Perfect score: 1596
Sequence: 1 MDSFETLLPESWLPKRLAS.....IAGSARRAHTROQFLGLS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89.5	5.6	1134	1 SREL_MOUSE	O9wtm3 mus musculus
2	87	5.5	457	1 NTRC_RHOCA	P09432 rhodobacter
3	86	5.4	672	1 TKT_RHOCA	O52723 rhodobacter
4	85	5.3	900	1 GDNH_CLOTM	P16218 clostridium
5	84.5	5.3	4303	1 PKD1_HUMAN	P98161 homo sapien
6	83.5	5.2	863	1 POL_IPHA	P04026 hamster int
7	83	5.2	285	1 ASPX_PAPHA	O06990 papio hamad
8	83	5.2	686	1 PIRB_ECOLI	P24355 escherichia
9	82.5	5.2	814	1 POL_IPMAI	P12894 mouse intra
10	82	5.1	510	1 MORE_MYCTU	O06220 mouse musculi
11	82	5.1	797	1 PKP3_MOUSE	O9qy23 mus musculi
12	81.5	5.1	453	1 RADA_PSEAE	P96963 pseudomonas
13	81.5	5.1	2564	1 SPCQ_HUMAN	O9h254 homo sapien
14	80.5	5.0	299	1 AMPM_METTH	O27355 metnanobact
15	80.5	5.0	391	1 ENTC_ECOLI	P10377 escherichia
16	80.5	5.0	392	1 PORL_THEMA	O05651 thermotoga
17	80.5	5.0	732	1 ERG7_HUMAN	P48449 homo sapien
18	80.5	5.0	884	1 YP67_MYCTU	O50654 mycobacteri
19	80	5.0	364	1 TFC1_AGRVI	O44471 agrobacteri
20	80	5.0	543	1 YJF8_YEAST	P31741 bacillus sp
21	79.5	5.0	718	1 CDG7_BACSI	P31747 bacillus sp
22	79.5	5.0	778	1 PCRA_MYCLE	O9c472 mycobacteri
23	79.5	5.0	859	1 OBP_HSVBC	P52377 bovine hearp
24	79.5	5.0	2212	1 RRP1_EBOZM	O05318 ebola virus
25	78.5	4.9	401	1 EAF3_YEAST	O12432 saccharomyc
26	78	4.9	295	1 MAUT_METEX	O49127 methyllobact
27	78	4.9	358	1 TTC3_AGRVI	O34296 agrobacteri
28	78	4.9	534	1 XYL1_ARATH	O91800 arabidopsis
29	78	4.9	558	1 P4HA_CAEEL	O10576 caenorhabdi
30	78	4.9	618	1 MBHL_RHOGE	P17532 rhodocyclu
31	78	4.9	707	1 GYS1_YEAST	P23337 saccharomyc
32	78	4.9	3164	1 POLG_TOMVU	P89309 t genome po
33	77	4.8	210	1 COT1_PHYCP	P41754 phytophthor

34	77	4.8	358	1 TTC5_AGRVI	O34295 agrobacteri
35	77	4.8	468	1 COBQ_THEAC	O9h128 thermoplas
36	77	4.8	538	1 NADB_PSEAE	O51363 pseudomonas
37	77	4.8	718	1 CDGF_BACCI	P30920 bacillus ci
38	77	4.8	1711	1 PRPO_RAP	O64612 rattus norv
39	76.5	4.8	777	1 ISOA_FLASP	O32611 flavobacter
40	76	4.8	265	1 ASPX_HUMAN	P26436 homo sapien
41	76	4.8	270	1 NUP3_PENQO	P24504 penicillium
42	76	4.8	316	1 FABH_HAEIN	P43711 haemophilus
43	76	4.8	562	1 HEX1_CANAL	P43077 candida alb
44	76	4.8	718	1 CDGF_BACLI	P14014 bacillus il
45	76	4.8	1696	1 PCK5_BRACL	O9nj15 branchiost

ALIGNMENTS

RESULT	1	STANDARD	PRT	1134 AA.
ID	SREL_MOUSE			
AC	O9WTN3			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Sterol regulatory element binding protein-1 (SREBP-1) (Sterol regulatory element-binding transcription factor 1).			
GN	SREBP1 OR SREBP1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE OF 1-41 FROM N.A. (ISOFORMS SREBP-1A AND SREBP-1C).			
RC	TISSUE=Liver;			
RX	MEDLINE=97216050; PubMed=9062340;			
RA	Shimomura I., Shimano H., Horton J.D., Goldstein J.L., Brown M.S.;			
RT	"Differential expression of exons 1a and 1c in mRNAs for sterol regulatory element binding protein-1 in human and mouse organs and cultured cells.";			
RL	J. Clin. Invest. 99:838-845(1997).			
RN	[2]			
RP	SEQUENCE OF 42-444 FROM N.A. (ISOFORMS SREBP-1A-W42 AND SREBP-1C-W42).			
RX	MEDLINE=99161303; PubMed=10052151;			
RA	Inoue J., Sato R.;			
RT	"A novel splicing isoform of mouse sterol regulatory element-binding protein-1 (SREBP-1).";			
RL	Biosci. Biotechnol. Biochem. 63:243-245(1999).			
RN	[3]			
RP	SEQUENCE OF 55-1134 FROM N.A.			
RA	Lloyd D.J., Shackleton S., Trembath R.C.;			
RT	"Mouse Srebp1.";			
RL	Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	RECONSTRUCTION FROM ESTS, AND CONCEPTUAL TRANSLATION.			
RA	Holo C.;			
RT	Unpublished observations (NOV-1999).			
CC	-1- FUNCTION: Transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) (5'-ATCACCAC-3'). Has dual sequence specificity, binding to both an E-box motif (ATACGGA) and to SRE-1 (5'-ATCACCAC-3'). Regulates the transcription of genes for sterol biosynthesis and the LDL receptor gene. SREBP-1A is much more active than SREBP-1C in stimulating transcription from SRE-1-containing promoters. SREBP-1C has a greater ability to stimulate transcription of genes involved in fatty acid metabolism as opposed to cholesterol metabolism.			
CC	-1- SUBUNIT: Forms a tight complex with SCAP in the ER membrane. Efficient DNA binding of the soluble transcription factor fragment requires dimerization with another bHLH protein.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein that moves from the endoplasmic reticulum to the Golgi in the absence of sterols.			
CC	-1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; SREBP-1A (SHOWN HERE), SREBP-1A-W42, SREBP-1C AND SREBP-1C-W42; ARE PRODUCED BY ALTERNATIVE SPLICING.			

CC -1- TISSUE SPECIFICITY: ISOFORM SREBP-1C PREDOMINATES IN LIVER
 CC ADRENAL GLAND, BRAIN AND ADIPOSE TISSUE, WHEREAS ISOFORM SREBP-1A
 CC PREDOMINATES IN SPLEEN. BOTH ISOFORMS ARE FOUND IN KIDNEY, THYMUS,
 CC TESTIS, MUSCLE, JEJUNUM, AND ILEUM.
 CC -1- PIM: Under sterol-depleted conditions, SREBPs are cleaved
 CC sequentially by site-1 and site-2 protease. The first cleavage
 CC occurs within the luminal loop and is directly regulated by
 CC sterols. The second cleavage by site-2 protease occurs within the
 CC first transmembrane stretch and liberates the soluble
 CC transcription factor. Cleavage by the cysteine proteases, caspase-
 CC 3 and caspase-7, is induced during apoptosis, independent of
 CC sterol levels.
 CC -1- MISCELLANEOUS: SREBPs have to be in a complex with the cleavage-
 CC activating protein (SCAP) to move to the Golgi and be cleaved by
 CC site-1 protease.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.jsb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL; AB017337; BAA74795.1; -
 CC EMBL; A1060765; -; NOT_ANNOTATED_CDS.
 CC EMBL; A1132428; -; NOT_ANNOTATED_CDS.
 CC EMBL; AA475250; -; NOT_ANNOTATED_CDS.
 CC EMBL; AA498625; -; NOT_ANNOTATED_CDS.
 CC EMBL; A1386259; -; NOT_ANNOTATED_CDS.
 CC EMBL; A1552467; -; NOT_ANNOTATED_CDS.
 CC EMBL; A1595781; -; NOT_ANNOTATED_CDS.
 CC EMBL; A1786503; -; NOT_ANNOTATED_CDS.
 CC EMBL; AF374266; AAK54762.1; -
 CC HSSP; P36956; IAm9.
 CC MED; MG1:107606; Stebfl.
 CC InterPro: IPR003015; HLH_Myc.
 CC InterPro: IPR001092; HLH_dlm.
 CC Pfam; PF00010; HLH; 1.
 CC SMART; SM00353; HLH; 1.
 CC PROSITE; PS00038; Helix_LOOP_HELIX; 1.
 CC Transcription regulation; Activator; DNA-binding; Lipid metabolism;
 CC Cholesterol metabolism; Nuclear protein; Transmembrane;
 CC Endoplasmic reticulum; Golgi stack; Alternative splicing;
 CC CYTOPLASMIC (POTENTIAL).
 CC POTENTIAL.
 CC TRANSMEM.
 CC DOMAIN 478 498
 CC TRANSMEM 499 536
 CC DOMAIN 537 557
 CC TRANSMEM 558 1134
 CC DOMAIN 1 60
 CC DVA_BIND 317 330
 CC DOMAIN 331 368
 CC DOMAIN 367 388
 CC DOMAIN 420 453
 CC SITE 451 452
 CC SITE 480 481
 CC SITE 519 520
 CC VARSPLIC 1 29
 CC VARSPLIC 90 131
 CC CONFLICT 272 276
 CC CONFLICT 510 510
 CC CONFLICT 613 613
 CC CONFLICT 621 621
 CC CONFLICT 625 625
 CC CONFLICT 633 633
 CC CONFLICT 906 906

SO SEQUENCE 1134 AA; 120505 MW; E0DF184F2A4FB7D0 CRC64;
 Query Match 5.6%; Score 89.5; DB 1; Length 1134;
 Best Local Similarity 21.3%; Pred. No. 6.3;
 Matches 74; Conservative 38; Mismatches 120; Indels 115; Gaps 17;
 OY 5 ETLPESEMLPRKPLASAKSGAYRHVTRORALELPYIEANPLWQSLYTRDASDA--- 61
 DB 774 ESLYSVACNPVYDLAQVYR--LFRHLLERALNC-IAQSPCAADG---DREFSDALGY 826
 OY 62 ---DMADLGLSPSPY-VSNMRYTTTG----- 85
 DB 827 LQLNCSDDAGAPACSESVSSSMATTPGVAKWASLTAAYVIMLRDEEAERLYP 886
 OY 86 ---HI-----YALKNPCLND--AARRPIYL-----LARV 112
 DB 887 LVEHLPQYLDTERPLRAHLXSFKAARALDHRVESPASALCEKASGTRSLAST 946
 OY 113 EOG-----LCDVL--GGDASYGHRITKNPLSTAHAFLMGPDALYELRALAHRTLD 160
 DB 947 PTGSSIDRAMOLLCDLLVARTSLMQROQSPASVQVAHGTSNGPOASALELRFQHDLS 1006
 OY 161 EIHALPEAGNP-----RNVTSYGRNVTLEDTRMAYRAVRHSMGCPVAE- 208
 DB 1007 SLRLAQSFRRPMRRVFLHEATARLMAGAPARTHQLD--RSLRRRAGSSGKGTTAEL 1064
 OY 209 -----W-EHTVEFHILNFTIIDEFATGP-IGLNELKHSLSR 248
 DB 1065 EPRPWRREHT---EALLLASCLLPAPFASBQGRMSMLAEARTYKE 1108
 RESULT 2
 NTRC_RHOCA STANDARD; PRT; 457 AA.
 AC P09432;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Nitrogen regulation protein ntrc.
 GN NTRC OR NTRF1.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 CC Rhodospirillum.
 OX NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SB1003;
 RX MEDLINE-89218961; PubMed-2710108;
 RA Jones R., Haselkorn R.;
 RT "The DNA sequence of the Rhodospirillum rubrum ntrc and ntrc
 RL gene analogues required for nitrogen fixation.";
 RL Mol. Gen. Genet. 215:507-516(1989).
 RP SEQUENCE FROM N.A.
 RC STRAIN-SB1003;
 RX MEDLINE-93360820; PubMed-8355615;
 RA Foster-Hartnett D., Cullen P.J., Gabbert K.K., Kranz R.G.;
 RT "Sequence, genetic, and lacZ fusion analyses of a ntrc3-ntrb-ntrc
 RL operon in Rhodospirillum rubrum.";
 RL Mol. Microbiol. 8:903-914(1993).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM NTRB/NTRC
 CC INVOLVED IN THE ACTIVATION OF NITROGEN ASSIMILATORY GENES SUCH AS
 CC GUNA. NTRC IS PHOSPHORYLATED BY NTRB AND INTERACTS WITH SIGMA-54.
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC INTERACTION ATP-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X12359; CAA30922.1; -
 DR EMBL: X72382; CAA51075.1; -
 DR PIR: S03321; S03321.
 DR HSP: P1028; LEV.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR002197; HTH_Fls.
 DR InterPro: IPR001789; Response_reg.
 DR InterPro: IPR002078; Sig54_interact.
 DR Pfam: PF02854; HTH_8; 1.
 DR Pfam: PF00072; response_reg; 1.
 DR Pfam: PF00158; Sigma54_activat; 1.
 DR SMART: SM00382; AAA; 1.
 DR SMART: SM00448; REC; 1.
 DR PROSITE: PS00110; RESPONSE_REGULATORY; 1.
 DR PROSITE: PS00675; SIGMA54_INTERACT_1; 1.
 DR PROSITE: PS00676; SIGMA54_INTERACT_2; FALSE_NEG.
 DR PROSITE: PS00688; SIGMA54_INTERACT_3; 1.
 DR PROSITE: PS00045; SIGMA54_INTERACT_4; 1.
 DR Nitrogen fixation; Transcription regulation; Repressor; Activator;
 KW DNA-binding; ATP-binding; Phosphorylation; Sensory transduction.
 FT DOMAIN 1 118
 FT MOD_RES 53 53 PHOSPHORYLATION (BY SIMILARITY).
 FT DOMAIN 113 343 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
 FT NP_BIND 171 178 ATP (POTENTIAL).
 FT DNA_BIND 425 444 H-T-H MOTIF (BY SIMILARITY).
 SO SEQUENCE 457 AA; 50042 MW; 7EB04BAFE4740306 CRC64;

Query Match 5.5%; Score 87; DB 1; Length 457;
 Best Local Similarity 24.6%; Pred. No. 3.3;
 Matches 75; Conservative 28; Mismatches 100; Indels 102; Gaps 16;

QY 5 EILFPPSMPLPRPLA--SAEKSQATYHNVTQRALSLPYT-----EANPLVMOSLYITDR 56
 DB 132 ETSRPG--TDLPVGRITAMQALYRLVARVMNADLPVIMGESGTGKSLIAKAI-----H 185
 QY 57 DASD-----ADMADLGLSPSYVSMNRV-----TTTGHYIY---ALK 92
 DB 186 DSDRTLEPFVYQAADLDGADGSPSLARRGRVLVPEVDGDDDETQGRIVRMADALP 245
 QY 93 NPV-----CLTDARRRRPIMLLARVEGGLCDVLGDAASYGHRITKKNPLSTAHAT 141
 DB 246 DPRIWATQVDLGLMEAGR-----REFQDLYYRLGG-----VT 280
 QY 142 LMGPADALYELALAHITDEIHAPL--AGNPRRNTSTVGNNVTLFDTTRMAYRAYR 199
 DB 281 LAVP-----ALREVEDIDPLAEHFLGAEERD-----GLGMRAFSAEANGLYRA-- 324
 QY 200 HSMGFPVAEMHTVFEHILHNET-----IIDEFATPGLI--NELKHLRSISIRW 249
 DB 325 YAMPGNVROLENTVRLVYVATSEETTRTEVEFVLGNQPAVEPLGAGGEGEKLSIAIRH 384
 QY 250 VWRNF 254
 DB 385 LRRYF 389
 RESULT 3
 TKT_RHOCA STANDARD; PRT; 672 AA.
 AC 052723; 068024;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transketolase (EC 2.2.1.1) (TK).
 GN TKTA.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 CC Rhodobacter.
 NCBI_TaxID=1061;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33303 / B10;
 RX MEDLINE=96186907; PubMed=8635754;
 RA Stry D'Aspremont R., Toussaint B., Vignais P.M.;
 RT Isolation of Rhodobacter capsulatus transketolase: cloning and
 RL sequencing of its structural tkla gene.";
 RN gene 169:81-84(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SB1003;
 RX MEDLINE=97404404; PubMed=9256491;
 RA Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Feinstein M.;
 RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter
 RT capsulatus SB1003".
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
 CC -1- CARBOLYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glucosealdehyde
 CC -3-phosphate = D-ribose 5-phosphate + D-xylulose 5-phosphate.
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.
 CC -1- PATHWAY: KEY ENZYME OF THE REDUCTIVE AND THE OXIDATIVE PENTOSE
 CC PHOSPHATE PATHWAYS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L4803; AAB06805.1; -
 DR EMBL: AF010496; AAC16110.1; -
 DR HSP: P23254; TTR.
 DR InterPro: IPR000360; Transketolase.
 DR Pfam: PF00456; transketolase; 1.
 DR Pfam: PF02780; transketolase; C; 1.
 DR Pfam: PF02779; transketolase; 1.
 DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE: PS00802; TRANSKETOLASE_2; 1.
 KW Transketolase; Thiamine pyrophosphate.
 FT CONFLICT 376 382 PESTIGS -> LNNHRL (IN REF. 1).
 FT CONFLICT 484 484 A -> T (IN REF. 1).
 SO SEQUENCE 672 AA; 72436 MW; B6B1E7DE1E66CA80 CRC64;

Query Match 5.4%; Score 86; DB 1; Length 672;
 Best Local Similarity 22.2%; Pred. No. 6.6;
 Matches 61; Conservative 37; Mismatches 103; Indels 74; Gaps 14;

QY 63 WAADLGLSPSYVSMNRV--TTTGHYIYVALKNVCLDARRRPINLLARVEGLCD--- 118
 DB 103 WGARMMG--HPEYGHLEGVETTTGPGGISTAVGMAYIAK-----SMAAFGRKLVDHKT 156
 QY 119 -VLGDASYGHRITKKNPLSTAHATLWGPADALYELALAHITDEIHAPLPEAGNRRNRYR 177
 DB 157 WVLADGCLMEGISTQEAIGLAG-----KQEDNLVIMDNNN----- 193
 QY 178 STYGRNVTLFDTTRMAYRAYRHSMGQVAEME-----HTVEHILNLN----- 221
 DB 194 ITIDGRVTVSDVTDOKARFA-----ASGMVLDSCDGHADIDIRALLTAARKAPVYL 245
 QY 222 ---ETII-----ADEPAT--GPIGLNELKHLRSISIRVMVNNF--PEFFRARKAISL 268
 DB 246 VDCKTILIGSPKAKADSYAVAGAPLGDALTK-LTRRAYGHEHGPVYIPATIKAEWALGA 304
 QY 269 RGASKGKGEHNGKIASGASRRRAHTRQOFLGLS 303
 DB 305 KGAARAEWEARLALPAG--KRAEFERQARQAVY 337
 RESULT 4
 GDNH_CLOTM

RA MEDLINE-97342914; PubMed-9199561;
RA Peral B., Gamble V., Strong C., Ong A.C.M., Sloane-Stanley J.,
RA Zervas K., Winearis C.G., Harris P.C.;
RT "Identification of mutations in the duplicated region of the
RT polycystic kidney disease 1 gene (PKD1) by a novel approach.";
RL Am. J. Hum. Genet. 60:1399-1410(1997).
RN [6]
RP VARIANT ALA-4058.
RX MEDLINE-97295081; PubMed-9150733;
RA Constantinides R., Xenophonos S.L., Neophytou P., Nomura S.,
RA Pierides A., Constantinou-Deltas C.D.;
RT "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
RT polycystic kidney disease 1 gene: evolution of alleles.";
RL Hum. Genet. 99:644-647(1997).
RN [7]
RP VARIANTS T-2760; P-2761; V-2763; T-2764; Q-2791; T-2826; L-3008 AND
RP L-3064.
RX MEDLINE-97449169; PubMed-9285784;
RA Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandolph M.A.,
RA Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;
RT "An unusual pattern of mutation in the duplicated portion of PKD1 is
RT revealed by use of a novel strategy for mutation detection.";
RL Hum. Mol. Genet. 6:1473-1481(1997).
RN [8]
RP VARIANT ADPKD THR-3678.
RX MEDLINE-97403939; PubMed-9259200;
RA Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,
RA Pignatti P.F.;
RT "Three novel mutations of the PKD1 gene in Italian families with
RT autosomal dominant polycystic kidney disease.";
RL Hum. Mutat. 10:164-167(1997).
RN [9]
RP VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
RX MEDLINE-98180892; PubMed-9521593;
RA Danielli C., Maheshwar M.M., Lazarou L., Davies F., Coles G.,
RA Ravine D., and recurrent mutations in the PKD1 (polycystic kidney
RT disease) gene.";
RL Hum. Genet. 102:216-220(1998).
RN [10]
RP VARIANT ADPKD MET-3375.
RX MEDLINE-99118881; PubMed-9921908;
RA Koptides M., Constantinides R., Kyriakides G., Hadjigavriel M.,
RA Patalis P.C., Pierides A., Deltas C.C.;
RT "Loss of heterozygosity in polycystic kidney disease with a missense
RT mutation in the repeated region of PKD1.";
RL Hum. Genet. 103:709-717(1998).
RN [11]
RP VARIANTS ADPKD L-324 AND S-845, AND VARIANTS R-1399 AND L-1786.
RX MEDLINE-99294580; PubMed-10364515;
RA Thomas R.L., McConnell R., Whitacker J., Kirkpatrick P., Bradley J.,
RA Sandford R.;
RT "Identification of mutations in the repeated part of the autosomal
RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range
RT PCR.";
RL Am. J. Hum. Genet. 65:39-49(1999).
RN [12]
RP VARIANTS ADPKD P-2392 AND F-2423, AND VARIANTS R-1399; Q-2548 AND
RP R-2638.
RX PubMed-10577909;
RA Watnick T., Phakdeekitcharoen B., Johnson A., Gandolph M., Wang M.,
RA Brietel G., Klinger R.W., Kimberling W., Gabow P., Germino G.G.;
RT "Mutation detection of PKD1 identifying a novel mutation common to
RT three families with aneurysms and/or very-early-onset disease.";
RL Am. J. Hum. Genet. 65:1561-1571(1999).
RN [13]
RP VARIANTS ADPKD L-3994--F-3996 DUP1, G-4136 AND C-4154, AND VARIANTS.
RX MEDLINE-20441957; PubMed-10987650;
RA Perichot R.A., Mercier B., Simon P.M., Whebe B., Clodes J., Ferec C.;
RT "DGC screening of PKD1 gene reveals novel mutations in a large cohort
RT of 146 unrelated patients";
RL Hum. Genet. 105:231-239(1999).
RN [14]
RP VARIANTS ADPKD 3748-R-V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.
RX MEDLINE-20112427; PubMed-10647901;
RA Aizal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,
RA Jeffery S.;
RT "Novel mutations in the 3' region of the polycystic kidney disease 1
RT (PKD1) gene.";
RL Hum. Genet. 105:648-653(1999).
RN [15]
RP VARIANTS ADPKD PRO-4225 AND TRP-4276.
RX MEDLINE-99217041; PubMed-10200984;
RA Badenas C., Torra R., San Millan J.L., Lucero L., Milla M.,
RA Estivill X., Darnell A.;
RT "Mutation analysis within the 3' region of the PKD1 gene.";
RL Kidney Int. 55:1225-1233(1999).
RN [16]
RP VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS CYS-2379; LEU-3066;
RP VAL-3139 AND LEU-3193.
RX MEDLINE-20311156; PubMed-10854095;
RA Perichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,
RA Clodes J., Ferec C.;
RT "Novel mutations in the duplicated region of PKD1 gene.";
RL Eur. J. Hum. Genet. 8:353-359(2000).
RN [17]
RP VARIANTS ADPKD PRO-2921 AND MET-3375, AND VARIANT LEU-3066.
RX MEDLINE-20382887; PubMed-10923040;
RA Koptides M., Mean R., Demetriou K., Constantinides R., Pierides A.,
RA Harris P.C., Deltas C.C.;
RT "Screening of the PKD1 duplicated region reveals multiple single
RT nucleotide polymorphisms and a de novo mutation in Hellenic
RT polycystic kidney disease families.";
RL Hum. Mutat. 16:176-176(2000).
RN [18]
RP VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.
RX MEDLINE-20514565; PubMed-11058904;
RA Aguiari G., Savelli S., Garbo M., Bozza A., Angello G., Penolazzi L.,
RA De Paoli Vitale E., La Torre C., Cappelli G., Piva R., del Senno L.;
RT "Novel splicing and missense mutations in autosomal dominant
RT polycystic kidney disease 1 (PKD1) gene: expression of mutated
RT genes.";
RL Hum. Mutat. 16:444-445(2000).
RN [19]
RP VARIANTS ADPKD SER-1166; GLU-1956; CYS-2408 AND GLY-2442--2443 INS,
RP AND VARIANTS HIS-1995 AND ASN-2604.
RX MEDLINE-20467506; PubMed-11012875;
RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,
RA Germino G.G.;
RT "Thirteen novel mutations of the replicated region of PKD1 in an Asian
RT population.";
RL Kidney Int. 58:1400-1412(2000).
RN [20]
RP VARIANTS ADPKD TRP-3753 AND ASN-3815.
RX MEDLINE-20275386; PubMed-10729710;
RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,
RA Hwang D.Y., Lee J.G., Namkoong Y., Lee C.C.;
RT "Novel mutations of the PKD1 gene in Korean patients with autosomal
RT dominant polycystic kidney disease.";
RL Mutat. Res. 433:39-45(2000).
RN [21]
RP VARIANTS ADPKD Q-13; F-75; C-139; 1992-F-1993 DELINS L; 2220-R--P-
RP 2224 DEL; D-2336; D-2752; 2762-TMR-2765 DUP1; M-2768; K-2771; P-2816;
RP S-2858; 3012-T--Y-3017 DEL AND 3748-L--R-3752 DEL, AND VARIANTS S-
RP 2674; M-2708; T-2734; L-2735; C-2765; M-2782; R-2814; G-2888; I-2905;
RP D-2966 AND L-3066.
RX MEDLINE-21063179; PubMed-11115377;
RA Rossetti S., Scimecki L., Gamble V., Burton S., Sneddon V., Peral B.,
RA Roy S., Bakraloglu A., Komel R., Winearis C.G., Harris P.C.;
RT "Mutation analysis of the entire PKD1 gene: genetic and diagnostic
RT implications.";
RL Am. J. Hum. Genet. 68:46-63(2001).
RN [22]
Query Match 5.3%; Score 84.5; DB 1; Length 4303;
Best Local Similarity 21.1%; Pred. No. 95;

Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps 9;

OY 4 FETTESWLPKPKPLASAKSGAVRHV-----TRQALPELYEANPLVQSLVI 53
DB 1391 FVQLEDDEAML-----VACAMPPEPRYRTWDFGEAAPTFRANGPEVTFEYRDP-----GSYLL 1443

OY 54 TDRQASDADNADLADGLSPSPVSNMRYTTHGHLYALKNPCLDAAARRRINLARVE 113
DB 1444 TVTASNNISAAANDSALVEQEPVLTSLKVNGLGLEQOPLYFASVGRPSY----- 1498

OY 114 QGLCDVLGDASVGHRTKNPLSTAHTL-----WGPADALYELRALAHTLD 160
DB 1499 --LMDLGGCMHEGPEVYHAINSTDFYVAGMNEVSSEMLVYKRRYRGV----- 1552

OY 161 EIHALPEAGNPRRYNSTVGNHTLFTPTRMAYRAVRHSN-----GGPVAEME 210
DB 1553 -----VNASRTVPLNGSVSFTSLGSDVRYSLDRCPTPIGCGPTISY- 1599

OY 211 HTVFHILNLTITADEFATGCPGLNFKLSRSISRMV 250
DB 1600 --TFRSVGTFTNIVTAE-----NEVGSADSIIFYV 1628

RESULT 6

POL_IPHA STANDARD; PRT; 863 AA.

AC P04026;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative POL polypeptide [Contains: Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
GN POL.
OS Hamster intracisternal A-particle (IAP-H18).
OC Viruses; Retroviral viruses; Retroviridae; Intracisternal A-particles.
OX NCBI_TaxID=11752;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85264989; PubMed=2991563;
RA Ono M., Toh H., Miyata T., Awaya T.;
RT "Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes."
RL J. Virol. 55:387-394(1985).
CC -1- MISCELLANEOUS: READTHROUGH OF THREE TERMINATORS MAY OCCUR: TAA BETWEEN CODONS ATT FOR 660-ILE AND AAA FOR 661-LYS, TAG BETWEEN CODONS TCC FOR 832-SER AND TAT FOR 833-TYR, AND TAG BETWEEN CODONS CCC FOR 859-PRO AND ATT FOR 860-ILE.
DR HSP, P03366; 1HMV.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR000477; RYase.
DR InterPro: IPR001584; Rye.
DR Pfam: PF00552; Integrase; 1.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF00075; RNaseH; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolyase; Nuclease;
KW Endonuclease; Polypeptide.
SQ SEQUENCE 863 AA; 97036 MW; 704AEF7C6B1D886 CRC64;

Query Match

Best Local Similarity 5.28; Score 83.5; DB 1; Length 863;
Matches 69; Conservative 44; Mismatches 110; Indels 121; Gaps 19;

OY 23 KSGAVRHVTRQALPELYEANPLVQSLVITDRDASDADNADLADGLSPSPVSNMRYT 82
DB 420 KTGAGAVYVKRVRISKVNRNSPOVVECLIVLE-----VLEAFPGP-----LNIVS 465

OY 83 TTHGHVIALKNPVCLTDA-----RRPILNARVEQGLVDGDSYGHRTKNPL-- 135
DB 466 DSSYV-----NAVNLLEIAGIRSSSRVANIPOKIOAALN-----RFPVEI 509

OY 136 --STHATLWGP-----ADALYELRL-----AHTDEIHALPEAGNPRRYTFRSYG 181
DB 510 THVRHSGILPGWISLGNLDLADKATLVATALSTHQAQKEFH-----KRHYAETLR 562

OY 182 RNVTL-----FDTTRWAV-RAVRHSWGPPVAEMHTV-----F 214
DB 563 RFAFSRKEAREITYQCQNCCEFLTPHMGINPRIR-----PLQMGMDYHIPSFGRL 617

OY 215 EIHILNLTITADEFATGCPGLNFKLSRSISRM-VWRN-----FTPETFR--A 261
DB 618 QYHVSVDTCSGVMFATPLTGKASYVIOHCLEAWSAMGPKRIKTDNGPAYTSQKRFQC 677

OY 262 ROKAISLNGASKGKEGKHGKI-----ASGSRRAH--TROQFL 299
DB 678 ROMDVT-----HLTGLPYNPOGQIGIVERAHRHTLSYL 709

RESULT 7

ASPX_PAPHA STANDARD; PRT; 285 AA.

AC P06990;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Acrosomal protein SP-10 precursor (acrosomal vesicle protein-1).
GN ACV1.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93183503; PubMed=8442953;
RA Freemanman A.J., Wright R.M., Flickinger C.J., Herr J.C.;
RT "Cloning and sequencing of baboon and cynomolgus monkey intra-acrosomal protein SP-10: homology with human SP-10 and a mouse sperm antigen (MSA-63)."
RL Mol. Reprod. Dev. 34:140-148(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94220578; PubMed=8167233;
RA Freemanman A.J., Wright R.M., Flickinger C.J., Herr J.C.;
RT "Tissue specificity of the acrosomal protein SP-10: a contraceptive vaccine candidate molecule."
RL Biol. Reprod. 50:615-621(1994).
CC -1- SUBCELLULAR LOCATION: NASCENT ACROSOMAL VESICLE OF GOLGI PHASE SPERMATID.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC -----

DR EMBL; S56458; AAB25692.1; -;
DR EMBL; S56458; AAB25693.1; -;
KW Signal; Glycoprotein; Repeat; Sperm; Alternative splicing.
FT SIGNAL 1 21
FT CHAIN 22 285
FT DOMAIN 66 95
FT REPEAT 66 70
FT REPEAT 71 75

POTENTIAL ACROSOMAL PROTEIN SP-10
3 X 5 AA REPEATS OF S-E-H-[GA]-A.
1-1.
1-2.

```

FT REPEAT 91 95 1-3.
FT DOMAIN 85 184 7 X 5 AA REPEATS OF S-G-E-H-[AL].
FT REPEAT 85 89 2-1.
FT REPEAT 120 124 2-2.
FT REPEAT 130 134 2-3.
FT REPEAT 145 149 2-4.
FT REPEAT 150 154 2-5.
FT REPEAT 170 174 2-6.
FT REPEAT 180 184 2-7.
FT DOMAIN 110 114 9 X 5 AA REPEATS OF [SV]-G-E-Q-[PSA].
FT REPEAT 110 114 3-1.
FT REPEAT 115 119 3-2.
FT REPEAT 125 129 3-3.
FT REPEAT 135 139 3-4.
FT REPEAT 150 154 3-5.
FT REPEAT 160 164 3-6.
FT REPEAT 165 169 3-7.
FT REPEAT 175 179 3-8.
FT REPEAT 190 194 3-9.
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC 172 205 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 285 AA: 30128 MW: 524E115F72AAD0E CRC64;

```

Query Match 5.2%; Score 83; DB 1; Length 285;
 Best Local Similarity 29.9%; Pred. No. 4.1;
 Matches 41; Conservative 11; Mismatches 55; Indels 30; Gaps 7;

```

OY 110 ARREGGLCVLGGDASVGRHRTKPNLSTAHATLWGPAD--ALVELNALHTLDEHIALPE 167
DB 16 ANGTQSQSSBSSG--SIDHOTSVOQLSGEFLSLENSDALEKLETLTSLSE-HGSSR 72
OY 168 AGNPRNRTSTVGRNVLTLETRMAVRAVRHSGPV----AEWEHTVFEH----- 216
DB 73 HSRREHTVAEHFGE-----HAESHAEGEPATGAEGEHTVGEQPSGRQPS 120
OY 217 -IHLNETIADFPANG 232
DB 121 GEHLSEGSISLG-EHAGS 136

RESUL 8
PTRB_ECOLI STANDARD; PRT; 686 AA.
ID PTRB_ECOLI
AC P24555; P78068;
DI 01-MAR-1992 (Rel. 21, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease II (EC 3.4.21.83) (Oligopeptidase B).
GN PTRB OR TLP OR B1845.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97251358; PubMed=9097040;
RA Kanatani A., Masuda T., Shimoda T., Misoka F., Lin X.S.,
RA Yoshimoto T., Tsuru D.;
RA "Protease II from Escherichia coli: sequencing and expression of the
RT enzyme gene and characterization of the expressed enzyme.";
RL J. Biochem. 110:315-320(1991).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97251358; PubMed=9097040;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF LYSYL
AND ARGININYL RESIDUES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of Arg-|-xaa and Lys-|-xaa bonds in
oligopeptides, even when the residue is proline.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 59A; ALSO KNOWN AS THE
PROLYL OLIGOPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
enterprises requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

Db	320	GETTETDMLVVEEQRGLTSLRQINRKTREVIAGIAFDPAVYTWLAIVNEPEFARLR	376
RESULT	9		
ID	POL_IPMA1	STANDARD:	PRT: 814 AA.
AC	P12894		
DT	01-OCT-1989	(Rel. 12, Created)	
DT	01-OCT-1989	(Rel. 12, Last sequence update)	
DT	01-NOV-1995	(Rel. 32, Last annotation update)	
DE	Probable POL polyprotein.		
OS	Mouse intracisternal A-particle (IAP-IL3).		
OC	Viruses; Retroid viruses; Retroviridae; Intracisternal A-particles.		
OX	NCBI_TaxID=11754;		
RN	SEQUENCE FROM N.A.		
RA	MEDLINE=86286596; PubMed=3016667;		
RT	Ymer S., Tucker W.Q.J., Campbell H.D., Young I.G.;		
RT	"Nucleotide sequence of the intracisternal A-particle genome inserted		
RL	5' to the interlukin-3 gene of the leukemia cell line WEHI-3B."		
CC	Nucleic Acids Res. 14:5901-5918(1986).		
CC	-1 MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-slb.ch/announce/ ,		
CC	or send an email to license@lsb-slb.ch).		
DR	EMBL; X04120; CAA27732.1; -.		
DR	PIR; A23597; GNMSIP.		
DR	HSSP; P04585; 1R1.		
DR	InterPro: IPR001037; Integrase_C.		
DR	InterPro: IPR003308; Integrase_zn.		
DR	InterPro: IPR002156; RNaseH.		
DR	InterPro: IPR00477; RYase.		
DR	InterPro: IPR001584; Rve.		
DR	Pfam; PF00552; Integrase_1.		
DR	Pfam; PF02022; Integrase_zn_1.		
DR	Pfam; PF00075; RNaseH_1.		
DR	Pfam; PF00665; rve_1.		
DR	Pfam; PF00078; rvc_1.		
KW	Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;		
KW	Endonuclease; Polyprotein.		
SC	SEQUENCE 814 AA; 91000 MW; A369620A50F729A CRC64;		
Query Match	5.2%; Score 82.5; DB 1; Length 814;		
Best Local Similarity	19.0%; Pred. NO. 17;		
Matches	54; Conservative 44; Mismatches 119; Indels 67; Gaps 11.		
Y	7 LPEPESMLPRKPL-----ASAERSGARHYTRQALALEPYEAMPVYQSLVIYD-----	55	
Db	363 VFPOITV-RNPLKNGIIVYTDGSKTGIGAYANGVSVKQYENSPQVECVLEVIKLT	421	
QY	56 -----RDAADADMAA-----DIAGLPSPSVSMKRVTTGHIYALKNPVCLTDAARR	104	
Db	422 FLKPLNTIVSDSYVYVNAVNLLEVAGVITPSSVYVAIIFOIQLVLLSRKSPVYITH-----	476	
QY	105 PINLLARVEQGLCDVLDGADSYGHRITRNPPLSTAHATIMGPADALYEURLAHTLDEIHA	164	
Db	477 -----VRAHSG-----LPPMALGNDLADKATKVVAALSSPVEAARNFH-----	521	
QY	165 LPEAGNPRRANTYTRSVGNRVYLFDTTRMA-----YRAVRSHSGGPFVAEMHTV---	213	
Db	522 TAEITLRSPSTLRKEARDIVITOCSCCEFLPYPHGVINPGRIR-----PLQVQMDVTHV	576	
QY	214 -----FEHIIILNFIITADFEA-----TGPLGLINELKHLSSISRW	249	
Db	577 SSFGKQLYIHWSIDPCSGIMFPASPLTGEKASVAVIHCHCEAMSAW	620	

RESULT	10			
ID	MURF_MYCTU	STANDARD:	PRT:	510 AA.
AC	006220;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	UDP-N-acetylmutaromylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)			
DE	(D-alanyl-D-alanine-adding enzyme).			
GN	MURF OR RV2157C OR MT2216 OR MTCY270.11.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RL	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland R., Eisen J.A., Carpenter L., White O.,			
RA	Petersen J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ernolova M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,			
RA	Bisbal W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: INVOLVED IN CELL WALL FORMATION CATALYSES THE FINAL			
CC	STEP IN THE SYNTHESIS OF UDP-N-ACETYLURAMUOYL-PENTAPEPTIDE, THE			
CC	PRECURSOR OF MUREIN (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: ATP + UDP-N-acetyluramoyl-L-alanyl-D-			
CC	glutamy-meso-2,6-diaminopimelate + D-alanyl-D-alanine = ADP			
CC	+ phosphate + UDP-N-acetyluramoyl-L-alanyl-D-glutamyl-6-			
CC	carboxy-L-lysyl-D-alanyl-D-alanine.			
CC	-1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.			
CC	-1- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	-1- SIMILARITY: BELONGS TO THE MURCODE FAMILY.			
CC				
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; Z59388; CAB08670.1; -			
DR	EMBL; AE007068; AAK46500.1; -			
DR	HSSP; P11880; IGC4.			
DR	TIGR; MT2216; -			
DR	TubercuList; RV2157C; -			
DR	InterPro; IPR000713; Mur_ligase.			
DR	InterPro; IPR004101; Mur_ligase.C.			
DR	Pfam; PF02875; Mur_ligase; 1.			
DR	PF02875; Mur_ligase.C; 1.			
DR	Peptidoglycan synthesis; Cell division; Cell wall; Ligase;			

KM ATP-binding; Complete proteome.
 FT NP_BIND 136 142 ATP (POTENTIAL).
 FT CONFIDENT 49 49 A -> R (IN REF. 2).
 FT CONFLICT 442 442 G -> E (IN REF. 2).
 SQ SEQUENCE 510 AA; 51632 MW; 7BEB3EDD1BA2BC4 CRC64;

Query Match 5.1%; Score 82; DB 1; Length 510;
 Best Local Similarity 26.1%; Pred. No. 10;
 Matches 80; Conservative 24; Mismatches 119; Indels 84; Gaps 19;

QY 28 RHVTRQALAEPLTEANPLMOSLYITDRASADMAADLAGLPSPEYSGMRTYTTGHI 87
 DB 224 REVIAQTKALP--QAVP--HSGAVVLN--ADDPAAVAAAKTLTAARVRSR--DNTGDV 275
 QY 88 VVALKNPVCTLDARRR-----PINLLARVEQGLCDVLDGASVYGRITKKNPLSTAHATLW 143
 DB 276 ---WAGVSLDELARPRFTLHAHQAEVRLGVC---GD---HOVT--NALCAAAVAL- 322
 QY 144 GPADALYELRALAHVTLDELTHALPEAGNPRNRVSTVGRNVTLFDTRMMAY----- 195
 DB 323 -ECGASVEQVAALT-----AAPVSRHRMQVT--TRGDVTVIDD-----AYNANPDSMR 370
 QY 196 -----RAVHSMG--GPVAMEHTVEFEHILNLTITADEFATGLINELK 240
 DB 371 AGIQAIAWIAHOPATRRSMALVGEAE-----LGEADIAEHDRIGRLAVR-- 416
 QY 241 HLSRSISRWVNRNFTPEFPAROKAISLRGASKGKGEKHKG-----GIASGASRAHT 294
 DB 417 ---LDVSRLLVVG--TGNSISAMHGAVALLEGASGSEATADHGADRTAVNVADGALALL 472
 QY 295 RQGFLEG 301
 DB 473 RAELRFG 479

RESULT 11
 PKP3_MOUSE
 ID PKP3_MOUSE STANDARD; PRT; 797 AA.

AC 090Y23;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Plakophilin 3.
 GN PKP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99310799; PubMed=10381383;
 RA Bonne S., van Hengel J., Nollet F., Kools P., van Roy F.,
 RT "Plakophilin-3, a novel armadillo-like protein present in nuclei and
 RT desmosomes of epithelial cells.";
 RL J. Cell Sci. 112:2265-2276(1999).
 CC -1- FUNCTION: MAY PLAY A ROLE IN JUNCTIONAL PLAQUES.
 CC -1- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes.
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 8 ARM REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF136719; AAD55892.1; -
 DR MGD: MGI:1891830; Pkp3.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.

DR SMART; SK00185; ARM; 3.
 DR PROSITE; PS50176; ARM_REPEAT; 1.
 KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
 KW Repeat.
 FT REPEAT 305 348 ARM 1.
 FT REPEAT 351 390 ARM 2.
 FT REPEAT 393 432 ARM 3.
 FT REPEAT 449 487 ARM 4.
 FT REPEAT 491 536 ARM 5.
 FT REPEAT 596 637 ARM 6.
 FT REPEAT 645 684 ARM 7.
 FT REPEAT 689 730 ARM 8.
 SQ SEQUENCE 797 AA; 87298 MW; A9DB0A76C79C126 CRC64;

Query Match 5.1%; Score 82; DB 1; Length 797;
 Best Local Similarity 22.9%; Pred. No. 18;
 Matches 69; Conservative 36; Mismatches 124; Indels 72; Gaps 16;

QY 9 PESMLPRKPLASAEKSGAYRHVTRQALAEPLTEANP-----LYWQSL----- 51
 DB 153 PTPMPTRPVSEFHRGGAASRADYD--TLSPSLRLGEGGLDRYSVSEGLEPAAASTYR 211
 QY 52 -VITDRASDA-----DNADLAGLPS-----PSVSNMRVTT-----TGHVYAL 91
 DB 212 AYAVKROASSGSSRAGGLDWEATEGPPRTIRAPAMKTLQRFOSHSRSGTGSVSGAG 271
 QY 92 KNPVCLDAAARRRPTINLLARVEG--LCDVLDGASVYGRITKKNPLSTAHATLMGPADALY 150
 DB 272 LEPAVAPRSVRSLSL---ADSGHLPDVRGLDSTYGR--TLQRLSSGFDDIDIPSAVKY 327
 QY 151 -----ELRALAHVTLDELTHALPEAGNPRNRVSTVGRNVTLFDTRMMAYRAVHSMG 204
 DB 328 LMSDPNLQYIGAAYIGHRCSDAAAKQAASLQAVRLVLFHNAG---EYGRHATGA 384
 QY 205 -----PVAMEHTVEFEHILNLTITADEF--ATGPL-GLNELKHSNIS 247
 DB 385 MRNLIDNVNKLALVE--ENGIFELRLTLRQ--DDELKKNVTGILMNLSSDHLKRLA 441
 QY 248 R 248
 DB 442 R 442

RESULT 12

RADA_PSEAE
 ID RADA_PSEAE STANDARD; PRT; 453 AA.

AC P96963;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA repair protein rada homolog (DNA repair protein sms homolog).
 GN RADA OR SMS OR PA4609.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRD1;
 RA Howell M.L., Heur M., Klotz M.G., Hassett D.J.;
 RT "Pseudomonas aeruginosa oxidative stress operon";
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN 12

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.U., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REPAIR OF ENDOGENOUS ACRYLATION
 CC DAMAGE (BY SIMILARITY).
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE RADA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U89384; AAB9466.1; -
 DR EMBL; AE004875; AAG07997.1; -
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR001984; Lon_endopep.
 DR InterPro; IPR001553; RecA.
 DR PRINTS; PR00830; ENDOLAPTASE.
 DR SMART; SM00382; AAA.1.
 DR PROSITE; PS50162; RECA.2; 1.
 KW DNA repeat; ATP-binding; zinc-finger; DNA-binding; Complete proteome.
 FT ZN_FING 10 27 C4-TYPE (POTENTIAL).
 FT NP_BIND 96 103 ATP (POTENTIAL).
 FT CONFLICT 142 103 E -> K (IN REF. 1).
 FT CONFLICT 240 240 E -> D (IN REF. 1).
 FT CONFLICT 243 244 GR -> AA (IN REF. 1).
 FT CONFLICT 247 247 L -> M (IN REF. 1).
 FT CONFLICT 258 258 V -> F (IN REF. 1).
 FT SEQUENCE 453 AA; 48411 MW; AAD23FEFAIEED5 CRC64;
 SO
 Query Match 5.1%; Score 81.5; DB 1; Length 453;
 Best Local Similarity 22.1%; Pred. No. 10;
 Matches 62; Conservative 32; Mismatches 91; Indels 95; Gaps 13;
 QY 1 MDSEFTLPESWLPKRP-----LASAEKSGA---YHVTROALPLEPYEA 43
 Db 170 IDSIQITFEQ-IGSAPGVAQVRESAAMLVYKASGSAFLVGHVYKREGAL-----A 222
 QY 44 NPLVMSQSLVTD-RDASDADMAADLAGLPSPSYSMNRVTTHGLVYALK--NPVCLTD 99
 Db 223 GPRVLEHWDVLYFEESDGRLLRLRAVKNRFGAVNVEGVGMDKGKEVNSPSAI-- 280
 QY 100 AARRPIMLLARVEQGLDVLGDSYGHRTKNPLSTAHATLMGPADALYELRALHTL 159
 Db 281 -----FLTRAGPAAV-----PGSYVMATWESRPMLEVEVALVDT- 314
 QY 160 DEIHALLPENGPRNRVTSYGRNVTL-EDTTRMAYRAVRHSHGCPVAEWEHTVEEHT- 217
 Db 315 -----SHLAMPRR-----VTGLGDQRLMLLAVLHRHG-IPTYQDQVFLANV 357
 QY 218 -----HLNFTIADFERATGPLGIN 237
 Db 358 GGVKYLETASDIALMAAVMSLRNRPDLPLDVLGEVGLS 397
 RESULT 13
 SPCQ_HUMAN STANDARD; PRT; 2564 AA.
 ID SPCQ_HUMAN Q9H254; Q9H254; Q9H254; Q9H254; Q9H254; Q9H254;
 AC Q9H254; Q9H254; Q9H254; Q9H254; Q9H254; Q9H254;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)
 DE (beta-IV spectrin).
 GN SPTBN4 OR SPTBN3 OR KIAA1642.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=21316449; PubMed=11294830;
 RA Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L.,
 RA Gwynn B., Peters L.L., Lux S.E.;
 RT "A new spectrin, beta-IV, has a major truncated isoform that
 RT associates with promyelocytic leukemia protein nuclear bodies and the
 RT nuclear matrix."
 RL J. Biol. Chem. 276:23974-23985(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4).
 RX MEDLINE=20539976; PubMed=11086001;
 RA Beggs S., Aguiar D., Dirks R. Jr., Maksimova E., Stabach P.,
 RA Hermet J.-M., Zhang J.-P., Philbrick W., Stepien V., Ort T.,
 RA Solimena M.;
 RT "Betatry spectrin, a new spectrin localized at axon initial segments
 RT and nodes of Ranvier in the central and peripheral nervous system."
 RL J. Cell Biol. 151:985-1002(2000).
 RN [3]
 RP SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).
 RX TISSUE=Brain;
 RC MEDLINE=20450683; PubMed=10997877;
 RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.
 RT XVIII. The complete sequences of 100 new cDNA clones from brain which
 RL code for large proteins in vitro."
 CC DNA Res. 7:273-281(2000).
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; are
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic
 CC islets.
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF311855; AAG2473.1; -
 DR EMBL; AF082075; AAG38874.1; -
 DR EMBL; AY004226; AAF93171.1; -
 DR EMBL; AY004226; AAF93172.1; -
 DR EMBL; AY004227; AAF93173.1; -
 DR EMBL; AB046862; BAB13468.1; -
 DR MIM; 606214; -
 DR InterPro; IPR001589; Actinin act. bind.
 DR InterPro; IPR001715; Calponin_hom.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001605; Spectrin_PH.
 DR Pfam; PF00307; CH. 2.
 DR Pfam; PF00169; PH. 1.
 DR Pfam; PF00435; spectrin. 18.
 DR PRINTS; PRO0683; SPECTRINPH.
 DR SMART; SM00033; CH. 2.
 DR SMART; SM00233; PH. 1.
 DR SMART; SM00150; SPEC. 16.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS50021; CH. 2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
 KW Alternative splicing.
 FT DOMAIN 1 282 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 61 165 CH 1.
 FT DOMAIN 180 282 CH 2.
 FT REPEAT 309 354 SPECTRIN 1.


```

FT REPEAT 398 419 SPECTRIN 2.
FT REPEAT 429 533 SPECTRIN 3.
FT REPEAT 535 642 SPECTRIN 4.
FT REPEAT 644 771 SPECTRIN 5.
FT REPEAT 773 879 SPECTRIN 6.
FT REPEAT 881 985 SPECTRIN 7.
FT REPEAT 1019 1086 SPECTRIN 8.
FT REPEAT 1088 1197 SPECTRIN 9.
FT REPEAT 1199 1303 SPECTRIN 10.
FT REPEAT 1305 1408 SPECTRIN 11.
FT REPEAT 1410 1513 SPECTRIN 12.
FT REPEAT 1515 1619 SPECTRIN 13.
FT REPEAT 1621 1725 SPECTRIN 14.
FT REPEAT 1727 1832 SPECTRIN 15.
FT REPEAT 1834 1940 SPECTRIN 16.
FT REPEAT 1942 2046 SPECTRIN 17.
FT REPEAT 2048 2107 SPECTRIN 18.
FT DOMAIN 2418 2527 PH.
FT VARSPLIC 1 1257 MISSING (IN ISOFORM 3).
FT VARSPLIC 1258 1286 AVQAEGLRSGNIGGEQAQAVTLLER -> MPHYPSCS
FT VARSPLIC 1287 1309 SAPSLGTPPQIQLEAHNR (IN ISOFORM 3).
FT VARSPLIC 1310 2564 NOENLRAGQOMOKLHDOLEOH -> CLIHPLALHPWE
FT VARSPLIC 2113 2154 PPIYLRSS (IN ISOFORM 2).
FT VARSPLIC 2155 2564 MISSING (IN ISOFORM 2).
FT VARSPLIC 2156 2564 TERKAEQSKPPPTPLGKRFEDPELAKAKAPLIRPGCY
FT VARSPLIC 2157 2564 E -> PRREDHNPQVQDPWOHTKPSLPKPRANEKTAR
FT VARSPLIC 2158 2564 RDGTCL (IN ISOFORM 4).
FT VARSPLIC 2159 2564 MISSING (IN ISOFORM 4).
FT VARSPLIC 2160 2564 MISSING (IN REF. 2).
FT VARSPLIC 2161 2564 L -> S (IN REF. 2).
FT VARSPLIC 2162 2564 E -> K (IN REF. 2).
FT VARSPLIC 2163 2564 E -> K (IN REF. 2).
FT VARSPLIC 2164 2564 G -> S (IN REF. 1).
FT VARSPLIC 2165 2564 G -> S (IN REF. 1).
SQ SEQUENCE 2564 AA; 288982 MW; 52CDE7D11D601ECC CRC64;

```

Query Match 5.1%; Score 81.5; DB 1; Length 2564;
 Best Local Similarity 22.3%; Pred. No. 91;
 Matches 65; Conservative 38; Mismatches 113; Indels 75; Gaps 14;

```

QY 11 SWL-PKRPPLASAEKSG-----AYRHYTORALFLPYEIANPLWMSQVITDRDASDMDA 64
DB 1529 AWQOEQLPLAMOTERNGICLAOVQOHKKNQGLRREIQAGSPRIEEYV-----ERA 1578
QY 65 ADLAGLPSPSYVSMNEVTTHGIVALKNPVCLTDARRRP--INLLARYEGCLDVLGG 122
DB 1579 GALASLRSPAEAVRGLQGLQSAWA-----GLREAEARQVLDAAFYQYFFDVAEV 1633
QY 123 DASYGRIT-----KNPLST-----AHATLMGPADALYELRALAHITDEIHALPEAG 169
DB 1634 EAWLGEQELLMSEDKGKDEQSTLQKLKHLQLEQGVEN-YE-ESTIQLSRCCRLLEMG 1691
QY 170 NP-----RRNVTRSTVGRNVTLFDTRRMAYRAVHSMGCVPAEHEHYV--- 213
DB 1692 HPDSEISRROQVRLYVALKELGEBRRVALRQOQWYVLSRO-----VSELEHWTAEK 1746
QY 214 -----FEHILNETIIDEFA--TGPIGLNELKHLSSISRWV 250
DB 1747 EYVAGSPELGDFEHVSVLOERF--SEFASSETMAGSERLAANQWVDELI 1795
RESULT 14
AMPK_METH
ID AMPK_METH STANDARD; PRT; 299 AA.
AC 027355;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Methionine aminopeptidase (EC 3.4.11.18) (MAP) (Peptidase M).
GN MAP OR MTH1296.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.

```

```

OX NCBI_TaxID=145262;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H:
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE
CC + PEPTIDE.
CC -1- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE
CC MAP FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000894; AAB85776.1; -.
DR HSSP; P56218; 1XGS.
DR MEROPS; M24.001; -.
DR InterPro; IPR002468; MAP_2.
DR InterPro; IPR001714; Methamino_Prase.
DR InterPro; IPR000994; Peptidase_M24.
DR Pfam; PF00557; Peptidase_M24; 1.
DR PRINTS; PR00599; MAPEPTIDASE.
DR PROSITE; PS01202; MAP_2; 1.
KW Hydroxylase; Aminopeptidase; Cobalt; Complete proteome.
FT METAL 84 84 COBALT 2 (BY SIMILARITY).
FT METAL 95 95 COBALT 1 AND 2 (BY SIMILARITY).
FT METAL 158 158 COBALT 1 (BY SIMILARITY).
FT METAL 191 191 COBALT 1 (BY SIMILARITY).
FT METAL 285 285 COBALT 1 AND 2 (BY SIMILARITY).
SQ SEQUENCE 299 AA; 33431 MW; DA37006A0638167 CRC64;

```

Query Match 5.0%; Score 80.5; DB 1; Length 289;
 Best Local Similarity 27.2%; Pred. No. 7.2;
 Matches 25; Conservative 9; Mismatches 33; Indels 25; Gaps 3;

```

QY 130 ITNPSTAHATLMGPADALYELRALAHITDEIHALPEAGN-----RRNVTRSTVG 181
DB 213 LRSPRLVHAR-----RVLCRKIRREYHALPFAQRMLEEYEDAKRLNASMR--- 258
QY 182 RNVTLFDTRRMAYRAVHSMGCVPAEHEHYV 213
DB 259 ---MLIOSRAIYPYHLREKSGANVAVQWEHTV 287
RESULT 15
ENTC_ECOLI
ID ENTC_ECOLI STANDARD; PRT; 391 AA.
AC P10377; P77099;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-OCT-1994 (rel. 30, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Isochorismate synthase entC (EC 5.4.99.6).
GN ENTC OR B0593 OR Z0735 OR ECS0632.
OS Escherichia coli, and
OS Escherichia coli O157:H7.

```

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE=90236256; PubMed=2110093;
 RT Elkins M.F., Earhart C.E.;
 RL "Optic factor from group A streptococci is an apoproteinase.";
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE=89123153; PubMed=2536681;
 RT Ozenberger B.A., Brickman T.J., McIntosh M.A.;
 RL "Nucleotide sequence of Escherichia coli isochorismate synthetase
 gene and evolutionary relationship of isochorismate synthetase
 and other chorismate-utilizing enzymes.";
 RL J. Bacteriol. 171:775-783(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12.";
 RN [4]
 RP Science 277:1453-1474(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 Federapfel N., Hyman R., Kaiman S., Komp C., Kuri O., Lew H., Lin D.,
 Nath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21079353; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobbeck E.J., Davis N.W., Lin A., Dimalanta E.T., Potamocis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RN [6]
 RP Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21136231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Iida T., Takami H., Honda T., Sasaki S., Ogasawara N., Yasunaga T.,
 Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RL "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genetic comparison with a laboratory strain K-12.";
 RN [7]
 RP DNA Res. 8:11-22(2001).
 RN [8]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=90230305; PubMed=2139473;
 RA Brickman T.J., Ozenberger B.A., McIntosh M.A.;
 RL "Regulation of divergent transcription from the iron-responsive fepB-
 entC promoter-operator regions in Escherichia coli.";
 RN [8]
 RP J. Mol. Biol. 212:669-682(1990).
 RN [9]
 RP PARTIAL SEQUENCE AND CHARACTERIZATION.
 RX MEDLINE=90241936; PubMed=2139795;
 RA Lin J., Quinn N., Berchold G.A., Walsh C.T.;
 RL "Overexpression, purification, and characterization of isochorismate
 synthase (EntC), the first enzyme involved in the biosynthesis of
 enterobactin from chorismate.";
 RL Biochemistry 29:1417-1425(1990).

CC -1- CATALYTIC ACTIVITY: Chorismate = Isochorismate.
 CC -1- PATHWAY: ENTEROBACTIN BIOSYNTHESIS. ENTEROBACTIN IS AN IRON-
 CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE
 CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.
 CC -1- SUBUNIT: MONOMER.
 CC -1- INDUCTION: By iron starvation.
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
 CC TRPE AND PABB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M24142; AA16100.1; -;
 DR EMBL; M36700; AA18491.1; -;
 DR EMBL; AE000165; AAC73694.1; -;
 DR EMBL; U82598; AAB40793.1; ALT_INIT.
 DR EMBL; AE005239; AAG54928.1; -;
 DR EMBL; AP002552; BAB34055.1; -;
 DR EMBL; X53274; CAA37371.1; -;
 DR PIR; J70497; SYECIK.
 DR PIR; S04323; S04323.
 DR ECGene; EG10261; entC.
 DR InterPro; IPR000350; Chorismate_bind.
 DR Pfam; PF00425; chorismate_bind.1.
 DR ProDom; PD000779; Chorismate_bind.1.
 DR Enterobactin biosynthesis; Iron transport; Isomerase;
 KW Complete proteome.
 FT CONFLICT 305 306 SG -> TA (IN REF. 1).
 SQ SEQUENCE 391 AA; 42931 MW; 62882569DFC1AC4 CRC64;

Query Match 5.0%; Score 80.5; DB 1; Length 391;
 Best Local Similarity 22.2%; Pred. NO. 10;
 Matches 61; Conservative 36; Mismatches 89; Indels 89; Gaps 13;

OY	9	PESWLPKPLASAEKSGRYHTROALELPYEANP-----LVMGSIYTTDDASD- 60
DB	91	PESW---QSFSGQEKQASARRRFTNSQSLNVEKRAIPQGTTFEOMVANAALATATPQYDK 147
OY	61	-----ADMAADLAGLPSPSYVMNRVYTTGHIYA-----LK----- 92
DB	148	VYLSRLIDITDDAID-SGVLLERLIAQNPVSYNHVPLADGVLGLASPELLLRKDER 206
OY	93	-NPVCLTDAARRPINLAR-----VEGGLCDVLCGDASGHRITKN 133
DB	207	FSSISLAGSARQPDVEDRENGNRLASFKDRHEHELVTOQMKFVLRERSSELT-VPSS 265
OY	134	PLSTAHATLM-----PADALYELBALAHIDEIHLPEAGNRRVNTSTGVRNTL 186
DB	266	POLITTPILMLATPEFGKANSQENALTLACLHPTPAL--SGFPQAAATQ-VIALELP 321
OY	187	EDTTRMAYRAVRHSGSPV-----AEWEHTV 213
DB	322	FD-----RELFGIVGWCDESGNGEWYTI 346

Search completed: October 4, 2002, 15:36:31
 Job time: 203 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 46.38 Seconds
(without alignments)
1130.174 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596
Sequence: 1 MDSFETLFPESWLPKRPLAS.....IAGSARRAHTROOFLGLS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:archaea:19:***
2: SP:archaea:***
3: SP:fungi:***
4: SP:human:***
5: SP:invertebrate:***
6: SP:mammal:***
7: SP:mhc:***
8: SP:organelle:***
9: SP:phage:***
10: SP:plant:***
11: SP:rodent:***
12: SP:virus:***
13: SP:vertebrate:***
14: SP:unclassified:***
15: SP:tvirus:***
16: SP:bacteriophage:***
17: SP:archaea:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1596	100.0	303	2	09F6J9	09F6J9 propionibac
2	403.5	25.3	295	2	09F5N6	09F5N6 rhodococcus
3	343	21.5	341	2	09ZEX6	09ZEX6 thiodacillu
4	338.5	21.2	310	2	030683	030683 rhodococcus
5	329	20.6	336	2	052148	052148 mycobacteri
6	322.5	20.2	308	2	051635	051635 shigella so
7	321.5	20.1	311	2	051631	051631 escherichia
8	320.5	20.1	308	2	051636	051636 shigella so
9	316.5	19.8	308	2	051639	051639 escherichia
10	311.5	19.5	306	2	09L8J6	09L8J6 rhodococcus
11	306	19.2	302	2	09L8J6	09L8J6 propionibac
12	304	19.0	310	2	045279	045279 brevibacter
13	303.5	19.0	311	2	051632	051632 escherichia
14	303	19.0	311	2	051630	051630 escherichia
15	298	18.7	297	2	09L8E8	09L8E8 plesiomonas
16	289.5	18.1	304	2	051637	051637 escherichia

17	286.5	18.0	304	2	051638	051638 escherichia
18	283	17.7	313	2	09K124	09K124 edwardsiella
19	282.5	17.7	297	2	051629	051629 shigella sp
20	282.5	17.7	304	2	051633	051633 escherichia
21	276.5	17.3	317	2	P97024	P97024 brevibacter
22	275.5	17.3	296	2	051634	051634 shigella so
23	268.5	16.8	320	2	09F4V6	09F4V6 brevibacter
24	266	16.7	296	2	09XBNI	09XBNI rhodococcus
25	259.5	16.3	306	2	045286	045286 bifidobacte
26	243.5	15.3	289	2	050999	050999 nelsisseria g
27	198	12.4	247	2	P94637	P94637 corynebacte
28	198	12.4	437	2	050474	050474 pseudomonas
29	191.5	12.0	437	2	09Z3S9	09Z3S9 pseudomonas
30	189.5	11.9	278	2	09X743	09X743 halomonas e
31	189	11.8	437	2	093TF3	093TF3 pseudomonas
32	180	11.3	320	2	056299	056299 thiodacillu
33	146	9.1	157	2	031107	031107 pseudomonas
34	117	7.3	524	2	0938U2	0938U2 corynebacte
35	108.5	6.8	524	2	093TF4	093TF4 corynebacte
36	107.5	6.7	521	2	093MK3	093MK3 corynebacte
37	105	6.6	84	2	046074	046074 corynebacte
38	99.5	6.2	2164	2	093N86	093N86 streptomyce
39	98	6.1	387	11	099JU4	099JU4 mus musculu
40	98	6.1	693	11	093JG9	093JG9 mus musculu
41	98	6.1	1461	16	005819	005819 mycobacteri
42	97	6.1	332	5	09VEF8	09VEF8 drosophila
43	93.5	5.9	228	17	09YD65	09YD65 aeropyrum p
44	93	5.8	355	12	09IGZ5	09IGZ5 maize strea
45	92	5.8	355	12	091MG6	091MG6 maize strea

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	303 AA.
09F6J9	09F6J9		
AC	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	REP1.		
OS	Propionibacterium freudenreichii.		
OG	Plasmid p545.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;		
OC	Propionibacterium.		
OX	NCBI_TaxID=1744;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-LMG16545;		
RX	MEDLINE-21091935; PubMed-1157209;		
RA	Jore J.P.M., van Luitk N., Luiten R.G.M., van der Werf M.J.,		
RA	Pouwels P.H.;		
RT	"Efficient transformation system for Propionibacterium freudenreichii		
RT	based on a novel vector.";		
RL	Appl. Environ. Microbiol. 67:499-503(2001).		
DR	EMBL; AF291751; MAG25294.1; -.		
DR	InterPro: IPR004322; Replicase.		
DR	Pfam: PF03090; Replicase; 1.		
KW	Plasmid.		
SQ	SEQUENCE 303 AA; 33613 MW; 5E6503BFA70A1AD7 CRC64;		

Query Match 100.0%; Score 1596; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 8.1e-133;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDSFETLFPESWLPKRPLASGAYRHTYRRALELPYENANPLVMOSLVTDDASD 60
DB	1	MDSFETLFPESWLPKRPLASGAYRHTYRRALELPYENANPLVMOSLVTDDASD 60
QY	61	ADMAADLAGLSPFSYVSMNRVTTGHIYVALKNPVCLTDAARRRPINLARVEGLCDVL 120

```

Db 61 ADMADLAGLSPSVSNRRITTTGHIYALKNPCLDARRRINILARVEGLCVL 120
Qy 121 GGDASYGHRITKPNLSTAHATLWGPADALYELRALAHTLDETHALPEAGNRRNVTRSTV 180
Db 121 GGDASYGHRITKPNLSTAHATLWGPADALYELRALAHTLDETHALPEAGNRRNVTRSTV 180
Qy 181 GRNTLTPTTTHMAYRAVRHSMGCVPAWEMHTVEPHILNETTIADDFATGPGILNKL 240
Db 181 GRNTLTPTTTHMAYRAVRHSMGCVPAWEMHTVEPHILNETTIADDFATGPGILNKL 240
Qy 241 HLSSISHWNRNFTPTFRAROKAISLRGASKGEGHKGIGASGASRRHTROGLE 300
Db 241 HLSSISHWNRNFTPTFRAROKAISLRGASKGEGHKGIGASGASRRHTROGLE 300
Qy 301 GLS 303
Db 301 GLS 303

```

```

RESULT 2
Q9F5N6 PRELIMINARY; PRT; 295 AA.
ID Q9F5N6; AC Q9F5N6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE REPLICATION PROTEIN.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RA Ryabchenko L.E., Novikov A.D., Golyshin P.N., Yansenko A.S.;
RT "Rhodococcus erythropolis plasmid pN30 putative replicase (ORF1), and
RT putative DNA-binding replication protein (ORF2).";
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF312210; AAC29854.1; -.
KW Plasmid.
SQ SEQUENCE 295 AA; 33678 MW; 05E9197C2F6DBB6 CRC64;

```

```

Query Match 25.3%; Score 403.5; DB 2; Length 295;
Best Local Similarity 36.0%; Pred. No. 1.4e-27;
Matches 104; Conservative 49; Mismatches 101; Indels 35; Gaps 11;

```

```

Qy 13 LPRKPLASAE-KSGAYRHTRORALELPYEANPLVMQSLVITPRDASADMAA--DLAG 69
Db 20 LPHRPFATNNLQROQYR-MSRDALAMRYVESHSHALLGSLVYICDHDVDAAMRAFEORSD 78
Qy 70 LPSPSYSVMNRVTGTHIVVAL-KNPVCLTDARRRPIINLARVEGLCVLGSDASYG 128
Db 79 HPANWVAOS-PSGRAHIGWLGPNHVCRTSDARLTPLRYAHRITETGKISVGGDFAVG 137
Qy 129 RITKNPLSTAHATLWGPADALYELRALA--HTLDETHALPEAGNRRNVTRSTVGRNTL 186
Db 138 QLTNPNLHPDWEETLYGPATP-YTLRQLATITTPROM-----PRRDRAVGLGRNTM 188
Qy 187 FDTTMMAYRAVRHSMGCVPAWEMHTVEPHILNETTIADDFATGPGILNKLHLSRSI 246
Db 189 FDATRRRAYRPMQWHRNCTGDMHVLVQCHAVN-----TEFTT-PLPTEVRAVTAOSI 242
Qy 247 SRWVRNFTPTFRAROKAISLRGASKGEGHKGIGASGASRRHTROGLE 295
Db 243 SKWIMRNTEEOYRARQAHL-----GQKGGKATTLAQEAVR 279

```

```

RESULT 3
Q9ZEX6 PRELIMINARY; PRT; 341 AA.
ID Q9ZEX6; AC Q9ZEX6;

```

```

DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE REPA PROTEIN.
GN REPA.
OS Thibacillus sp.
OC Bacteria; Proteobacteria; beta subdivision; Thiodacillus.
OX NCBI_TaxID=924;
RN [1]
RP SEQUENCE FROM N.A.
RA Perera J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ007958; CAA07792.1; -.
DR InterPro; IPR004322; Replicase.
DR Pfam; PF03090; Replicase; 1.
KW Plasmid.
SQ SEQUENCE 341 AA; 38283 MW; E0F1B8A7532F55AD CRC64;

```

```

Query Match 21.5%; Score 343; DB 2; Length 341;
Best Local Similarity 32.1%; Pred. No. 3.6e-22;
Matches 98; Conservative 40; Mismatches 129; Indels 38; Gaps 11;

```

```

Qy 13 LPRKPLASAEKSGAYRHTRORALELPYEANPLVMQSLVITPRD--ASDADW--AADL 67
Db 19 LPHRPFCTDIDLTLAGLQISLKTALQRAYIQNHQMIWAMYVDVDRACVNPDMWPIWES 78
Qy 68 AGLPSPYSVMNRVTGTHIVVALKNPCLTDARRRPIINLARVEGLCVLGSDASYG 127
Db 79 VGLPPNPNAVNNHRRKRLHLLYLEAGVCRTLEHLKRLRYIASERYTAVLADPQYA 138
Qy 128 HRTKNPLSTAHATLWGPADALYELRALAHTLDETHA---LPEAGNRRNVTRSTVGR 182
Db 139 GLICKNPENDR---W---QVWEIHGNPYTLAEIAAYVDLTAKAARLPSNASESFGIGR 190
Qy 183 NVTLEDTTMMAYRAVRHSMG-GVPAWEMHTVEPHILNETTIADDFATGPGILNKLH 241
Db 191 NCTMPTDRFMAVYGAVERVYMPNGIDRWSKAVLEQLQAIN---GEPPO-PLPFAEVKA 244
Qy 242 LRSISRWVRNFTPTFRAROKAISLRGASKGEGHKGIGASGASRR 291
Db 245 TGKISGWTWRNMP---AGLDLIERHTPEKQAEGRKATNQADIAVMGIGASGAR 301
Qy 292 AHTRQ 296
Db 302 LSSSQ 306

```

```

RESULT 4
Q30683 PRELIMINARY; PRT; 310 AA.
ID Q30683; AC Q30683;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE REPLICATION PROTEIN.
GN REPA.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-NI86/21;
RC MEDLINE=98015402; PubMed=9353918;
RA De Mot R., Nagy I., De Schrijver A., Pattanapipitpaisal P.,
RA Schoofs G., Vanderleyden J.;
RT "Structural analysis of the 6-kb cryptic plasmid pFAJ2600 from
RT Rhodococcus erythropolis NI86/21 and construction of Escherichia coli-
RL Microbiology 143:3137-3147(1997).

```



```
Db 179 LFEWTRKWAYRAIRGWM--PVFSQWLDVAVIQREVMYVNSLPV-----PLSPACRAIGK 230
OY 245 SISRWVWNPETPEFR-----AROKAISLRGASKGKGGHK---GGIASGA 288
Db 231 SIATKTHNFPETPEFAQYVADHTHPEIQANGRKGGKIGAKSKRGAVATSA 282

RESULT 7
ID 051631 PRELIMINARY; PRT; 311 AA.
AC 051631.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE REP PROTEIN.
OS Escherichia coli.
OC Plasmid COL2-GEI602.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven ColE2-related
   plasmids.";
RL J. Bacteriol. 176:7233-7243(1994).
DR EMBL; D30057; BAA06295.1; -.
DR InterPro; IPR000792; HTH LuxR.
DR InterPro; IPR004322; Replicase.
DR InterPro; IPR001440; TPR.
DR Pfam; PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 311 AA; 35081 MW; F91655855E321C2A CRC64;

Query Match 20.1%; Score 321.5; DB 2; Length 311;
Best Local Similarity 31.3%; Pred. No. 2.5e-20;
Matches 94; Conservative 53; Mismatches 124; Indels 29; Gaps 12;

OY 10 ESWLPRKPLASAEKSGAVRHVTRORALEPYTEANPLVMSQSLV--TDRDASDADWADL 67
Db 9 EENLPHRYHTDGLAFGLRISGKRALARIQONOPHAQFWLVVDVREGAIDWSDRN 68
OY 68 AGLPSPSYVMNRVTGTHIYVALKNPVCLTPDAARRPINLARVEQGLCDVIGSDASYG 127
Db 69 A--PAPNITVKMPVNGHAHLVLTALNIAVTPADASVKALKYAAIAERALCEKLGADVYNS 126
OY 128 HRTKNPSTAHATIMGADALYELRALAHITLDEL-HALPEGNRRNVYTS-TVGRNVT 185
Db 127 GLICKNPHELE---W---LVMEREBAVTLDELADYLDLSASARRSIDKHVYGRNRY 178
OY 186 LFDTRMAVAYRAVHSMGCPV-AEWEHTVFHEIHLNETIADFEATPGLINELKHHSR 244
Db 179 LFEWTRKWAYRAIRGWM--PVFSQWLDVAVIQREVMYVNSLPV-----PLSPACRAIGK 230
OY 245 SISRWVWNPETPEFRAROKAISLRGASKGKGGHKGGIASGA---SRAATRQOFLG 301
Db 231 SIATKTHNFPETPEFAQYVADHTHPEIQANGRKGGKIGAKSKRGAVATSA 282

RESULT 8
ID 051636 PRELIMINARY; PRT; 308 AA.
AC 051636.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE REP PROTEIN.
OS Shigella sonnei.
OC Plasmid COL6-C714.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
```

```
OX NCBI_TaxID=624;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven ColE2-related
   plasmids.";
RL J. Bacteriol. 176:7233-7243(1994).
DR EMBL; D30061; BAA06299.1; -.
DR InterPro; IPR004322; Replicase.
DR InterPro; IPR001440; TPR.
DR Pfam; PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 308 AA; 34922 MW; 55863D1650C03EAE CRC64;

Query Match 20.1%; Score 320.5; DB 2; Length 308;
Best Local Similarity 32.2%; Pred. No. 3e-20;
Matches 94; Conservative 45; Mismatches 122; Indels 31; Gaps 11;

OY 10 ESWLPRKPLASAEKSGAVRHVTRORALEPYTEANPLVMSQSLV--TDRDASDADWADL 67
Db 9 EENLPHRYHTDGLAFGLRISGKRALARIQONOPHAQFWLVVDVREGAIDWSDRN 68
OY 68 AGLPSPSYVMNRVTGTHIYVALKNPVCLTPDAARRPINLARVEQGLCDVIGSDASYG 127
Db 69 A--PAPNITVKMPVNGHAHLVLTALNIAVTPADASVKALKYAAIAERALCEKLGADVYNS 126
OY 128 HRTKNPSTAHATIMGADALYELRALAHITLDEL-HALPEGNRRNVYTS-TVGRNVT 185
Db 127 GLICKNPHELE---W---LVMEREBAVTLDELADYLDLSASARRSIDKHVYGRNRY 178
OY 186 LFDTRMAVAYRAVHSMGCPV-AEWEHTVFHEIHLNETIADFEATPGLINELKHHSR 244
Db 179 LFEWTRKWAYRAIRGWM--PVFSQWLDVAVIQREVMYVNSLPV-----PLSPACRAIGK 230
OY 245 SISRWVWNPETPEFR-----AROKAISLRGASKGKGGHK---GGIASGA 288
Db 231 SIATKTHNFPETPEFAQYVADHTHPEIQANGRKGGKIGAKSKRGAVATSA 282

RESULT 9
ID 051639 PRELIMINARY; PRT; 308 AA.
AC 051639.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE REP PROTEIN.
OS Escherichia coli.
OC Plasmid COL6-J.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven ColE2-related
   plasmids.";
RL J. Bacteriol. 176:7233-7243(1994).
DR EMBL; D30064; BAA06302.1; -.
DR InterPro; IPR004322; Replicase.
DR InterPro; IPR001440; TPR.
DR Pfam; PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 308 AA; 34851 MW; 2245A0EA06FBF085 CRC64;

Query Match 19.8%; Score 316.5; DB 2; Length 308;
Best Local Similarity 30.9%; Pred. No. 6.9e-20;
Matches 90; Conservative 48; Mismatches 124; Indels 29; Gaps 10;
```


RESULT 12
ID Q45279 PRELIMINARY; PRT: 310 AA.
AC Q45279;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
OS Brevibacterium plasmid POTATIVE THEA REPLICASE (ORF310).
OC Brevibacterium plasmid.
CC Plasmid autonomous circular plasmid PBL1.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Micrococcales; Brevibacteriaceae; Brevibacterium.
OX NCBI_TaxID=1703;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RBL;
RX MEDLINE=97092385; PubMed=6938050;
RA Ankri S., Bouvier I., Reyes O., Predali F., Leblon G.;
RT "A Brevibacterium linens PBL1 replicon functional in Corynebacterium
glutamicum.";
RT Plasmid 36:36-41(1996).
RL EMBL: U39878; AAB03568.1; .
DR InterPro: IPR004322; Replicase.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 310 AA; 35266 MW; 30B4DF1F26A98BF4 CRC64;
Query Match 19.0%; Score 304; DB 2; Length 310;
Best Local Similarity 32.8%; Pred. No. 8.8e-19;
Matches 106; Conservative 34; Mismatches 137; Indels 46; Gaps 13;
QY 3 SPETFPESWLPKPKLASAE-KSGAYRHVTRORALELPYEANPLVMOSLVTDDASDA 61
DB 5 STET-WGGMWLPMPPLASDLDLGGYR-TSRNNALEKRIENPQSLNLVVDIDHDA 62
QY 62 -----DNAADLAGLPSPSYVSMNRVTGTHIVALKNEVCITDAARRRPINLLARV 112
DB 63 LMRAMNRRKAM-----QPNAAVVENPANGHAHVAWLAEPVTEYARRKPLAYAAV 114
QY 113 EOGLCDVVGDSYSGHRTKKNPLSTAHNTLWGPADALYELRALHTLDEINHLPRAG--- 169
DB 115 TEGLRSSVDGKGYSGLTITKNTHDQEWASW-LTDHLNLEBELTEHLVSDPMPPDSMOR 173
QY 170 NPRNRVSTVGKRVNLTEDTRMAVRAVRHSGPVAEWEHTVEHILNETI----- 225
DB 174 TKRKNPV--GGRNCTLEPYRMVDYVARPI---RKRNHPPEEDHDLAALVNLQ 227
QY 226 -ADEPATGPLGNEKLHSRISRKVWRNFT---PETFRAROKAISLRGASGSGEGH 280
DB 228 GGNSTFSEALPASEIRATIRSFYKWTITRYTGWLDRTTSQSKSAHYHNT---GRKGGL 284
QY 281 KGIASGASRAHTRQOFLGEGIS 303
DB 285 KGVVSGQVR---RQKALERVS 303
RESULT 13
ID Q51632 PRELIMINARY; PRT: 311 AA.
AC Q51632;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
OS Escherichia coli.
OC Plasmid COLE2Imm-K317.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven COLE2-related
plasmids.";
RT J. Bacteriol. 176:7233-7243(1994).
DR EMBL: D30058; BAA06296.1; .
DR InterPro: IPR000792; HTH_LuxR.
DR InterPro: IPR004322; Replicase.
DR InterPro: IPR001440; TPR.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 311 AA; 35370 MW; 5569BA824F2B2646 CRC64;
Query Match 19.0%; Score 303.5; DB 2; Length 311;
Best Local Similarity 30.3%; Pred. No. 9.8e-19;
Matches 89; Conservative 53; Mismatches 125; Indels 27; Gaps 10;
QY 10 ESWLPKPKLASAEKSGAYRHVTRORALELPYEANPLVMOSLVT--TDRDASDADMAADL 67
DB 9 EENLPHRPHHTDDLAFGRISGKGRALLARYIQNQPHAQEWLVDPVREGAALIDMSDRN 68
QY 68 AGLPSPSYVMNRVTGTHIVYALKNPVCLTDAARRRPINLLAREOGCLCVLGSDSYXG 127
DB 69 A--PAPNITVKNPVNGHAHLIALNIAVTRAPDASVKALKTAAYERSLCERLCAADVNS 126
QY 128 HRTKNPLSTAHNTLWGPADALYELRALHTLDEI-HALPENGPRNRVTRS-TVGRNVT 185
DB 127 GLICKNPFHLE---W---LVMEWREARYITDELADYLDLSASRRSIDKHGGRNCH 178
QY 186 LFDTRMAVRAVRHSGPVAEWEHTVEHILNETIADDEFATGPLGNEKLHSRIS 245
DB 179 LFEMTRKMAVRAIROGWPA-FSQWIDAVIQREMYNASLPV-----PLSPPECAIRKIS 231
QY 246 ISRWWRNFTETFEFAROKAISLRGASGSGEGHKGISGASGRATROFL 299
DB 232 IAKYTHRNFTPETF---AQYVADHTPEIQARKGKGSIAG---EAYVDRFM 279
RESULT 14
ID Q51630 PRELIMINARY; PRT: 311 AA.
AC Q51630;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
OS Escherichia coli.
OC Plasmid COLE2-CA42.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven COLE2-related
plasmids.";
RT J. Bacteriol. 176:7233-7243(1994).
DR EMBL: D30056; BAA06294.1; .
DR InterPro: IPR000792; HTH_LuxR.
DR InterPro: IPR004322; Replicase.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ SEQUENCE 311 AA; 35298 MW; F2E75BECA0DBA137 CRC64;
Query Match 19.0%; Score 303; DB 2; Length 311;
Best Local Similarity 30.4%; Pred. No. 1.1e-18;
Matches 90; Conservative 55; Mismatches 123; Indels 28; Gaps 12;
QY 8 FPESWLPKPKLASAEKSGAYRHVTRORALELPYEANPLVMOSLVT--TDRDASDADMAA 65

Db 8 FDEN-LPHRPHHTDGLAFLRISGKALLARYIQONQPHQFWLVPDVBECALIDW-S 65
 QY 66 DLGLSPSPSYVSMNRVTTHGHIVYALKNPVLCDARRRPINLARVEOGICDVLGDAS 125
 Db 66 DL-NAPAPNITIKNPVGNHLLYALNIAVTAADASVKALKYAAERLALCELTGADV 124
 QY 126 YGHRITKNPLSTAHATMGADALYELRALAHILDEI-HALPEAGNRRNVTNS-TVGRN 183
 Db 125 YSGLICKNPFHLE---W---QVMEWEKEAYTLDELADYLDLSTARSIDKHYGMGRN 176
 QY 184 VTLEDTRMAVAVRASHMSGVPAWEHTEVEHILNLTETIADFEATGPIGLNELKHL 243
 Db 177 CHLEFEMTKKAYRAIRKQWPA-FSQWLDAVIQRYEMTNASLPV-----PLSPPECRAIG 229
 QY 244 RSISRWWVNRFTPETFRAROKAISLRGASKGEGHKGIGASGASRRATROQFL 299
 Db 230 KSIKAKYTHRRFTJAEF---AQYVADHTPELOAKRGKRKGIAKG---EAYDDKRFM 279

RESULT 15
 Q9RLEB PRELIMINARY; PRT; 297 AA.
 ID Q9RLEB
 AC Q9RLEB;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE REPLICON PROTEIN.
 GN REP.
 OS Plesiomonas shigelloides (Aeromonas shigelloides).
 OC plasmid p1184.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Plesiomonas.
 OX NCBI_TaxID=703;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O36:H34;
 RX MEDLINE=21223402; PubMed=11322823;
 RA Avison M.B., Walsh T.R., Bennett P.M.;
 RT "PUB060: a broad host-range, DNA polymerase-I-independent ColE2-like
 RT plasmid."
 RL plasmid 45:88-100(2001).
 DR EMBL; AJ249644; CAB56518.1; -;
 DR InterPro; IPR004322; Replicase.
 DR Pfam; PF03090; Replicase; 1.
 KW plasmid.
 SQ SEQUENCE 297 AA; 33820 MW; FBDDEF09EEFBCE2 CRC64;

Query Match 18.7%; Score 298; DB 2; Length 297;
 Best Local Similarity 33.4%; Pred. No. 2.8e-18;
 Matches 101; Conservative 33; Mismatches 138; Indels 30; Gaps 12;

QY 6 TLPEESLPPKPLASAEKSGAIRVTRQRALELPYIEN-PLVWQSLVT-TDRDASDADW 63
 Db 7 TLFNDR-LPHKPYFSDLOQGVIRAGKERALLAKYIQONPHAMYWLCFDDVRAGAIDW 65
 QY 64 AADLAGLSPSPSYVSMNRVTTHGHIVYALKNPVCITDAARRPINLARVEOGICDVLGD 123
 Db 66 -ADL-GAPAPNITIKNPVGNHLLYALNIAVTAADASVKALKYAAERLALCELTGADV 123
 QY 124 ASYGHRTKNPLSTAH--ATLMGPADALYELRALAHILDEIHALPEAGNRRNVTSTVG 181
 Db 124 AGYSGLICRNP-NHLHMQITVWQP--ELYTLDMADYLD----LGAANDREILPDYGLG 175
 QY 182 RNTLTPTTRMAVAVRASHMSGVPAWEHTEVEHILNLTETIADFEATGPIGLNELKH 241
 Db 176 RNCLEFEDTKKAYRAIRKQW-PEYSQWLQACIERARAYNLQFSA-----PLDENFVWG 228
 QY 242 LSRSSIRHW-----WRNFTPETFRAROKAISLRGASKGEGHKGIGASGASRRAH 293
 Db 229 IAKSISKWMTVYTRSLGDFDEVYKILTHSPYQAYNGRRSKGGGRPSIGEPWLALGISRSY 288
 QY 294 TR 295

Db 289 FR 290

Search completed: October 4, 2002, 15:36:07
 Job time: 179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 20.69 Seconds
(without alignments)
357.707 Million cell updates/sec

Title: US-09-720-583A-2

Perfect score: 1596
Sequence: 1 MDSFETLPESWLPKPLAS.....IAGSARRAHTRQDFLEGIS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	5.3	4302	3	US-08-658-136-5
2	84.5	5.3	4303	2	US-08-460-751-2
3	83.5	5.2	450	4	US-09-191-879-2
4	83.5	5.2	1043	3	US-08-928-361B-30
5	83.5	5.2	1721	3	US-08-700-651-5
6	83.5	5.2	1721	3	US-08-928-361B-6
7	83	5.2	285	1	US-08-292-045-5
8	78.5	4.9	1745	2	US-09-031-485-33
9	78.5	4.9	1745	2	US-08-847-429A-33
10	78.5	4.9	1745	2	US-09-065-474-33
11	78	4.9	973	3	US-08-904-452-4
12	77	4.8	319	2	US-08-683-908-1
13	77	4.8	1711	2	US-08-342-930-2
14	77	4.8	2539	4	US-09-413-814-42
15	77	4.8	3739	4	US-09-320-878-2
16	77	4.8	3739	4	US-09-105-537-33
17	77	4.8	11877	4	US-09-105-537-6
18	76.5	4.8	746	1	US-08-476-519-11
19	76.5	4.8	746	5	PCT-US95-09323-11
20	76.5	4.8	777	1	US-08-476-519-2
21	76.5	4.8	777	5	PCT-US95-09323-2
22	76	4.8	265	1	US-08-292-045-2
23	76	4.8	285	1	US-08-292-045-10
24	75.5	4.7	685	3	US-08-947-965-74
25	75.5	4.7	501	3	US-08-906-791-2
26	75.5	4.7	501	4	US-09-111-730-1
27	75.5	4.7	773	3	US-08-564-264-1

28	75	4.7	250	4	US-09-216-295-14	Sequence 14, Appl
29	75	4.7	378	4	US-09-500-569-10	Sequence 10, Appl
30	75	4.7	867	2	US-08-938-365-2	Sequence 2, Appl
31	75	4.7	1042	3	US-08-928-361B-11	Sequence 11, Appl
32	75	4.7	1837	3	US-08-928-361B-5	Sequence 5, Appl
33	74.5	4.7	554	4	US-08-943-714-2	Sequence 2, Appl
34	74	4.6	285	1	US-08-292-045-7	Sequence 7, Appl
35	74	4.6	417	2	US-08-672-564-9	Sequence 9, Appl
36	74	4.6	580	2	US-08-672-564-1	Sequence 20, Appl
37	73.5	4.6	417	4	US-08-867-611-20	Sequence 25, Appl
38	73.5	4.6	417	5	PCT-US92-06655A-25	Sequence 9, Appl
39	72.5	4.5	177	2	US-08-737-825-9	Sequence 32, Appl
40	72.5	4.5	606	2	US-08-577-492-32	Sequence 32, Appl
41	72.5	4.5	606	4	US-09-079-630-32	Sequence 2, Appl
42	72.5	4.5	828	3	US-08-481-435-2	Sequence 10, Appl
43	72.5	4.5	850	3	US-08-481-435-10	Sequence 18, Appl
44	72.5	4.5	964	3	US-08-860-339-18	Sequence 5, Appl
45	72.5	4.5	1611	2	US-08-804-227C-5	

ALIGNMENTS

RESULT 1
US-08-658-136-5
; Sequence 5, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31, 845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-658-136-5

Query Match 53%, Score 84.5, DB 3, Length 4302;
Best Local Similarity 21.1%, Pred. No. 26;
Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps 9;

Qy	4	FETLEPSPMPPRRPLASAEKSGVRRH-----	-TORALEIPTYEANPLYWOSLVI	53
Dd	1391	FVLGDEAML-----	VACAMPPEPYRTTWDPGEBAAPTRARPEYTFILRDP-----	GSLIV 14433
Qy	54	TDRBASADMAADLAGICPSSTIYSAMRVTTTGHITYALKNPVCLDAAARRPINLARVE		113
Dd	1444	TVTASNNISANOSALAVEOEPALVTSIKVNSLGLELOOPLFSAVNGRGPRASY----		1498
Qy	114	OGLDLVDVGDSAGHYRTTKNPLSTAHTL-----	WGPADALYTELRALAHTLTD	160
Dd	1499	-LMDLDGGMLBSEPEVTHAYNSTGGPFYRVAGWNVEVSSEAMLVYKKRRGLV		1552
Qy	161	EIHALLPENGRNRNVTRSYGRNVLTLEDTRRMATYAARHSN-----	GGPVAEWE	210
Dd	1553	-----VNASTVYPLNGSVSFSTSEAGSDVRYSWVLCDRCPIPGCGPTIIS-		1599
Qy	211	HVVFEHIHLNETRIADREFATGPLGNELKHLSRSTRV		250
Dd	1600	-TFRSYGTENIIVTAI-----	NEGSADSIFEVY	1628

```

US-08-460-751-2
Sequence 2, Application US/08460751
Patent No. 5891628
GENERAL INFORMATION:
APPLICANT: Redders, Stephen
APPLICANT: Schneider, Michael
APPLICANT: Glucksman, Sandra
TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,751
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/413,580
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7638-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-460-751-2
Query Match 5.3%; Score 84.5; DB 2; Length 4303;
Best Local Similarity 21.1%; Pred. No. 26;
Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps

```

```

Db      1391 FVLQGDSEAML-----VACAMPEPPRYRTWDFGTEEAAPTRARGPEVTFIYRDP-----GSYL 1443
OY      54   TDRASADMAADLADGLPSPSYVSVMNRVYTTTGHVYALKNPCLDARARRPINLAAYE 113
Db      1444 TVTASNNISANDSALVAEYDEPVLVTSIKYNSLGLDEDPFLFSAVGRGRPAS----- 1498
OY      114 QGLCDVLGSDASFGHRTTKNPFLSTAHATL-----WGPAALYELRALAHITLD 160
Db      1499 ---LMDLDGGWGLGPEPTAHYNTSGDFTVRVAGMNEVSRSEAMLVNTVRRVGLV---- 1552
OY      161 EIHALPEAGNPRNRVNTKSTVGRNVLTLEDTTRMAYPAVHSH-----GGFVAEME 210
Db      1553 -----VANASTVYPLNGSVSFSSFLSEAGSDVRYSWLCLDRCPPIGGPTIY- 1599
OY      211 HTVEFHILNERTLIADFEFTGPLGLNELKHLHSRISIRVY 250
Db      1600 ---TRSYGTENIIVTAL-----NEVGSADSLTFVYV 1628

```

```

RESULT: 3
US-09-191-879-2
: Sequence 2, Application US/09191879
: Patent No. 6204014
: GENERAL INFORMATION:
: APPLICANT: May, Earl W.
: APPLICANT: Jaworski, Deborah D.
: APPLICANT: Wang, Min
: APPLICANT: Warren, Richard L.
: APPLICANT: Lennox, Anna L.
: APPLICANT: Ingraham, Karen A.
: APPLICANT: Ray, Jennifer
: APPLICANT: Chalker, Alison F.
: APPLICANT: Holmes, David J.
: APPLICANT: Mazulla, Marie J.
: TITLE OF INVENTION: DnaB
: FILE REFERENCE: GM10183
: CURRENT APPLICATION NUMBER: US/09/191,879
: CURRENT FILING DATE: 1998-11-13
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: FastSeq for Windows Version 3.0.
: SEQ ID NO 2
: LENGTH: 450
: TYPE: prt
: ORGANISM: Streptococcus pneumoniae
US-09-191-879-2

```

[illegible]

```
RESULT 4
US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIRSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928, 361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verdy, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480,76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-30

Query Match          5.2%; Score 83.5; DB 3; Length 1043;
Best Local Similarity 25.3%; Pred. No. 3,6;
Matches 74; Conservative 30; Mismatches 100; Indels 89; Gaps 17;

QY 52 VITDDASD-ADWADLADLPSPSYVSMNRVT-----TTGHIY----- 88
DB 452 VVDEAKDQADKDGKGLIVPTNSINKDPVTNTQYNTGNTGINDETKGVIPGSL 511
QY 89 -YALKRPVCLTDARRRPINLARVEQGLCDVLGDSYGHRI---TKNPLSTAHATLWG 144
DB 512 NVPSEFTPOQTEITGKPVDTVG-----LPYDPSGEIIDPATKLPISVA----- 559
QY 145 PADATYELRALHTDEIHALP---EAGNPRNRVTRSTVGRNVTLFTTRMANAYRAVRHS 201
DB 560 -GDEIL-TEVLNITDDEVGLPIDLETGLPRDVSGLPOLPNGTIVDPENK--KPIPS 614
QY 202 WGGPAWEHETVFEHILNLTIDEFATG-PL-----GLNELKHSRSISRWVWNRFP 256
DB 615 HSGFI---NGTSGEOSH-----EKDPSTGKPLDPNTGLHPDEDSGL-----INP 657
QY 257 ETRAROKAISLRGASKG-----GREGHGKGIASGASRRARTRROOFLGSL 303
DB 658 ET-----GDKLQSHSGTFMPVPGKPGENGIM-----TPQIILEALN 696

RESULT 5
```

```
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INJECTIONS
; FILE REFERENCE: 480,19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match          5.2%; Score 83.5; DB 3; Length 1721;
Best Local Similarity 25.3%; Pred. No. 8;
Matches 74; Conservative 30; Mismatches 100; Indels 89; Gaps 17;

QY 52 VITDRDASD-ADWADLADLPSPSYVSMNRVT-----TTGHIY----- 88
DB 1130 VVDEAKDQADKDGKGLIVPTNSINKDPVTNTQYNTGNTGINDETKGVIPGSL 1189
QY 89 -YALKRPVCLTDARRRPINLARVEQGLCDVLGDSYGHRI---TKNPLSTAHATLWG 144
DB 1190 NVPSEFTPOQTEITGKPVDTVG-----LPYDPSGEIIDPATKLPISVA----- 1237
QY 145 PADATYELRALHTDEIHALP---EAGNPRNRVTRSTVGRNVTLFTTRMANAYRAVRHS 201
DB 1238 -GDEIL-TEVLNITDDEVGLPIDLETGLPRDVSGLPOLPNGTIVDPENK--KPIPS 1292
QY 202 WGGPAWEHETVFEHILNLTIDEFATG-PL-----GLNELKHSRSISRWVWNRFP 256
DB 1293 HSGFI---NGTSGEOSH-----EKDPSTGKPLDPNTGLHPDEDSGL-----INP 1335
QY 257 ETRAROKAISLRGASKG-----GREGHGKGIASGASRRARTRROOFLGSL 303
DB 1336 ET-----GDKLQSHSGTFMPVPGKPGENGIM-----TPQIILEALN 1374

RESULT 6
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIRSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verryn, Hana
:
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
:
: INFORMATION FOR SEQ ID NO: 6:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1721 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-928-361B-6

```

Query Match	5.2%	Score 83.5	DB 3	Length 1721
Best Local Similarity	25.3%	Pred. No. 8		
Matches 74	Conservative 30	Mismatches 100	Indels 89	Gaps 17

QY	52	WITDRDASD-ADMAADLAGLPSPSYVSNMRT	-----	TTGTHV	-----	88
Db	1130	VVPDEEKDDQARKGKGLIVPPNINSINKDPVNTQY	-----	NTTGTGIIINIPETQKVI	-----	1189

QY 89 -YALKNPEVCLTDAAARRRPINLLARVEOGLDVLGGDASYGHR|--TKNPPLSTAHATLWG 144
|||
bb 1190 NPPSFNPQOGLTEITGKRVDTATG-----LPYDSTGEITDIPATKLPISVA----- 1237

OY 145 PADALYELRALAHNLDEIHALP---EAGNPRNRVTRSTGRNVTLFDITRMAYRAVRHS 201
-|-| | | | | | | | | | | | | | | |
DB 1238 -GDEFI--TEFVNITDDEVGTPIIDIFETGI PRDVSGCIBOIENETIMPSNK-----KPTICS 1292

09	202	WGSRAEMHTVEFHILNFIADPATG-PL---	GLNELKHLRSISRWYWRNPTP	256
09	1203	WGSRAEMHTVEFHILNFIADPATG-PL---	GLNELKHLRSISRWYWRNPTP	256

QY 257 ETEFRAROKAISLRGASKG-----GKEGCHKGCIASGASRRRAHTRODTLEGLS 303

20	1968	STANDARDIZED FIVE YEAR VOLUNTARY	1 FEBRUARY 1974
RESULT	7		

US-08-492-045-5
: Sequence 5, Application US/08392045
: Patent No. 5603005
: GENERAL INFORMATION:

APPLICANT: HERR, JOHN C.
APPLICANT: WRIGHT, RICHARD M.
TITLE OF INVENTION: PRIMATE INTRA-ACROSOMAL SPERM ANTIGEN
FIELD OF INVENTION: FOODS IN CONDOMS

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia

```

? COUNTRY : U.S.A.
? ZIP : 22202
? COMPUTER READABLE FORM :
? MEDIUM TYPE : Floppy disk

```

```

;
;
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; Copyright 1987 by PatentIn, Inc. All rights reserved.

```

[illegible]

```

? APPLICATION NUMBER: US/08/292,045
? FILING DATE:
?
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/318,551
? FILING DATE: 03-MAR-1989
? APPLICATION NUMBER: US 07/481,491
? FILING DATE: 16-FEB-1990
? APPLICATION NUMBER: US 07/858,798
? FILING DATE: 27-MAR-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Oblon, No. 5602005mad F.
? REGISTRATION NUMBER: 24,618
? REFERENCE/DOCKET NUMBER: 494-109-0 CIP
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 413-3000
? TELEFAX: (703) 413-2220
? TELEX: 248855 OPAT UR
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 285 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-292-045-5

```

Query Match	5.2%	Score 83	DB 1	Length 265
Best Local Similarity	29.9%	Pred. No. 0.54		
Matches 41	Conservative 11	Mismatches 55	Indels 30	Gaps 7

Qy 110 ARVEGGLCDVLGGDASGHRITKNPSTAHATLWGPAD--ALYELRLAHATLDEIHALPE 167
Db 16 ARGTSQSDSSSG--SIDHQTSVQQSLGSEFFSLNENPDAEALYETASGLTSLSE-HSSSE 72

Oy 168 AGNPRRVNSTGVGRNVLTDTTMMAYRAVRHSGWGPV---AEWEHTVEEH----- 216
| : |
73 HGSREHTVAEHPGE-----HAESHBASGSPAAATGAEGCHTTVGOPSPGOPS 120
| : |

OY	217	-IHLNETIAD	232
		:	:
DB	121	GEHSGEOSITG-FH	136

```
RESULT      8
US09-03106-33
```

```

: Sequence 33 Application US/09031485
: Patent No. 5824306
: GENERAL INFORMATION:

```

1 APPLICANT: Xing Wang
2 APPLICANT: Blehm, E. Scot
3 TITLE OF INVENTION: DIROFILARIA AND BROGIA ANKYRIN
4 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
5 TITLE OF INVENTION:

TITLE OF INVENTION: USES THEREOF
 NUMBER OF SEQUENCES: 85
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Carol Talkington Versar, Ph.D.

ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
COUNTY: Colorado

COUNTRY : USA
ZIP : 80525
COMPUTER READABLE FORM :

ADDRESS FILE: PROPER DATA
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

2. PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/847,429
FILING DATE: 24-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1745 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-031-485-33

Query Match 4.9%; Score 78.5; DB 2; Length 1745;
Best Local Similarity 22.8%; Pred. No. 30;
Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YIEANPLVMQSLVITDRDASD-----ADMAADLA 68
DB 1385 AITLPEYTGPEPWSKLFYSEASLTKYVGAFHETAEPDNPLAHVALLIGADWHRLAR 1444
QY 69 GLPSP-----SYVSNRVTTTGHIVYALK--NPVCLDPAARRPINILARVEOGL 116
DB 1445 ALEVPDIDROVRHOLVGLAVTILRIWIFLKEQATPALSALORIGRDVVR-EMDR 1503
QY 117 CVDLGGDASVGHRTKNPLSTAHATLWGPADALYELRALAHLDLDELTHALPEAGNPRRN-- 174
DB 1504 AEKLDG-----LEGTPTS--HIS--GPS-----ITLSSTLLEY--AGDRRRHAE 1541
QY 175 VTRS-----TVGNNTLFD-----TTRMAVRAVRHSM 202
DB 1542 VTMAOQRLAEPFPOQVNGTTPGDPEEPKEQSFHEEEVAVSEIRTVRTERHVDSE 1601
QY 203 GGPVAMEHTV---FEHILNLTIIADFPATGCLGLNELKHLRSISRWVNRFTPETF 259
DB 1602 NGPIVE-ERTITTYEDDVAVNEEIVDKIV--PLNEEQEKMDRVN-REVENNF----- 1652
QY 260 RAROKAISLRGA---SKGKGEGHKGGIASGASRRRAHTRQOFLEG 301
DB 1653 --EQETSKEGTFCQCTHEKEKDDGSLKTKMDSHVRQIFPDG 1696

RESULT 9

US-08-847-429A-33
Sequence 33, Application US/08847429A
Patent No. 5827692
GENERAL INFORMATION:
APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/847,429A

FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1745 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-847-429A-33

Query Match 4.9%; Score 78.5; DB 2; Length 1745;
Best Local Similarity 22.8%; Pred. No. 30;
Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YIEANPLVMQSLVITDRDASD-----ADMAADLA 68
DB 1385 AITLPEYTGPEPWSKLFYSEASLTKYVGAFHETAEPDNPLAHVALLIGADWHRLAR 1444
QY 69 GLPSP-----SYVSNRVTTTGHIVYALK--NPVCLDPAARRPINILARVEOGL 116
DB 1445 ALEVPDIDROVRHOLVGLAVTILRIWIFLKEQATPALSALORIGRDVVR-EMDR 1503
QY 117 CVDLGGDASVGHRTKNPLSTAHATLWGPADALYELRALAHLDLDELTHALPEAGNPRRN-- 174
DB 1504 AEKLDG-----LEGTPTS--HIS--GPS-----ITLSSTLLEY--AGDRRRHAE 1541
QY 175 VTRS-----TVGNNTLFD-----TTRMAVRAVRHSM 202
DB 1542 VTMAOQRLAEPFPOQVNGTTPGDPEEPKEQSFHEEEVAVSEIRTVRTERHVDSE 1601
QY 203 GGPVAMEHTV---FEHILNLTIIADFPATGCLGLNELKHLRSISRWVNRFTPETF 259
DB 1602 NGPIVE-ERTITTYEDDVAVNEEIVDKIV--PLNEEQEKMDRVN-REVENNF----- 1652
QY 260 RAROKAISLRGA---SKGKGEGHKGGIASGASRRRAHTRQOFLEG 301
DB 1653 --EQETSKEGTFCQCTHEKEKDDGSLKTKMDSHVRQIFPDG 1696

RESULT 10

US-09-065-474-33
Sequence 33, Application US/09065474
Patent No. 6063599
GENERAL INFORMATION:
APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 171
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/065,474

```

; FILING DATE: 24-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HM-5-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-065-474-33

Query Match      4.9%; Score 78.5; DB 3; Length 1745;
Best Local Similarity 22.8%; Pred. No. 30;
Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YTEANPLVMQSLVITDRDASD-----ADMADLA 68
   |||
Db 1385 AITLPETGEPWVSKRLFTSEASLTKEYVGAFHETAEPDNPLAHVALLIGADMHRLAR 1444
   |||
QY 69 GLPSP-----SYVSNMKNVTGTHIYALK---NPVCLTDARRRPINILARVEGL 116
   |||
Db 1445 ALEVPDIDIRVHQVGLVLEAVITLRIMFLKKEGATPVRLRSALQIRGDDYVR-EMDR 1503
   |||
QY 117 CDVIGGASGHRITKPLSTAHNTLWGPADALVELALHTLDEIHALPAGCPRRN-- 174
   |||
Db 1504 AEKIDG-----LEGTPVS--HIS--GPS-----ITLSTLLEV-----AGDRRRHAE 1541
   |||
QY 175 VTRS-----TVGRNVTLPD-----TRMAYRAVVRHSW 202
   |||
Db 1542 VTMAQOQLAGEPFGQVGYNGTCDPEEPKEQSHREEEVAVSEITVYNTENHVDSE 1601
   |||
QY 203 GGPVNEHTV---FEHILHNETIADEPATGFLGNETLHLSRSISRWVNRFTPEF 259
   |||
Db 1602 NGPIVE-ERTITTYEDDVAVNEEIVDKIV--PLNEEEOKKMRMV-REVENMF----- 1652
   |||
QY 260 RARKAISLGA-----SKGKEGKHGKGIASGASRAHTRQOFLG 301
   |||
Db 1653 --EQQTSKGTGCGQTHERKDDGSLKTYMKDSHVRIFFDG 1696

RESULT 11
US-08-904-452-4
; Sequence 4, Application US/08904452
; Patent No. 6083742
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo M.
; TITLE OF INVENTION: Mammalian Deep Orange Proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,452
; FILING DATE: 31-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
```

```

; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1240.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-904-452-4

Query Match      4.9%; Score 78; DB 3; Length 973;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 49; Conservative 27; Mismatches 106; Indels 44; Gaps 7;

QY 5 ETLPEEWLPKPLASAEKSGAVRHVTRORALEPLYTEANPLVMQSLVITDRDASDADWA 64
   |||
Db 19 QTCGSPVGIPIHSGVVAHLEKEVIFTKQRYDFPSPSRITSLVYSCNQLCKSGKDTLLR 78
   |||
QY 65 ADLAGLPSPTVSKNRYT-----TTGHIYVALKNPVCL---TDAARRRPINILA 110
   |||
Db 79 IDLGKASEPNRVELGRKDDAKVHKMFLDHTGSHLLVALSSTEVLYMNRNGOKARP---LA 135
   |||
QY 111 RVEQGLDVGADASYGHRITKNP-----LSTAHNTLWGPADALVELRALA 156
   |||
Db 136 RAKQQLVESVGMKAMGNESSTGPILVGTAQGQIFELALSLSEGLGPGADLT----- 189
   |||
QY 157 HTLDEIHALPEAGNP---RRNVTSTVGRNVTLEDTTRMAYRAV 198
   |||
Db 190 --FRPLVYLNEEGGPAPVCSLEAERGPDRGFEVY-ATTROQLFGFI 232
   |||

RESULT 12
US-08-683-908-1
; Sequence 1, Application US/08683908
; Patent No. 5830693
; GENERAL INFORMATION:
; APPLICANT: Shimizu, Sakayu
; TITLE OF INVENTION: A Regulatory Factor Involved in
; TITLE OF INVENTION: Expression of Nitrlase Gene, and its Gene.
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,908
; FILING DATE: 19-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185626/1995
; FILING DATE: 21-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Davidson, Clifford M
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 3821004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 768-3800
; TELEFAX: (212) 382-2124
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 319 AMINO ACID RESIDUES
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Rhodococcus rhodochrous
INDIVIDUAL ISOLATE: J1
CELL TYPE: unicellular organism
US-08-683-908-1

Query Match 4.8%; Score 77; DB 2; Length 319;
Best Local Similarity 21.3%; Pred. No. 3;
Matches 48; Conservative 33; Mismatches 94; Indels 50; Gaps 11;

QY 4 FETLPRESW-----LPKRLASAEKSGAVRHVTRORALEPLTEANPLVMOSLVITDRD 57
DB 114 YDLFTDMRMSVFOFPKRSFGFTD-----FGRMTAVVGGDRIGRVSSFMSTIND 168
QY 58 ASDADMAADLAGLPSPSYVSMNRTTGHIVYALKNPVCLTDARRRPINLLARVEOGLC 117
DB 169 ATDAGLAEEVAVLSNSAVDLLSAIRF-----ELADQAAASDGLL---CVLAVIRQNL 220
QY 118 D-VLGSDASGHRITKNPLSTAH-----ATLKPADALYELRALHTLDETHALPEAGNP 171
DB 221 DPNLCASQIAAEHNVS---VRLHRLFSANGGVAEHIRMIR-----LERIKT--ELADP 270
QY 172 RRVNRSYVGRNVTLFDTTRMAY-----RAVRHSGGPAVEM 209
DB 271 -----TSRRYTTISLARKKGFLLDPSTFSRAFKDYGITAREW 307

RESULT 13

US-08-342-930-2
; Sequence 2, Application US/08342930
; Patent No. 5821084

GENERAL INFORMATION:

APPLICANT: OLMSTED, ELIZABETH A.

APPLICANT: MAURO, LAURA J.

APPLICANT: DAVIS, ALAN R.

APPLICANT: DIXON, JACK E.

TITLE OF INVENTION: OSTROBLAST-TESTICULAR PROTEIN TYROSINE

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/342,930

FILING DATE: 21-NOV-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: KONSKI, ANTOINETTE F.

REGISTRATION NUMBER: 34,202

REFERENCE/DOCKET NUMBER: 20344-20975.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1711 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-342-930-2

Query Match 4.8%; Score 77; DB 2; Length 1711;
Best Local Similarity 20.1%; Pred. No. 42;
Matches 56; Conservative 39; Mismatches 109; Indels 74; Gaps 11;

QY 20 SAEKSGAVRHVTRORALEPLTEANPLVMOSLVITDDASDADMAADLAGLPSPSYSMN 79
DB 580 AALKASWYHPGGGRDAFHLRLRPLTSEKVLPREAONFSMAQLTACCEFOVOLS-- 637
QY 80 RVTTGHIYVYALKNPVCLTDARRRPINLLA---RVEOGLCDVGLGDASYGHRITKNPL 135
DB 638 --TLMGSESSSANANGWTPPSAPTLVNVTSDAPTQVSWAHVPGGRSRY--QVTLVYE 693
QY 136 STAHAT-LMGPADALYELRALHTLDETHALPEAGNRRRVNYSYGRNVTLFDTTRMA 194
DB 694 STRTATSIMGP-----KEDGTSFLGLTRGT-----R 719
QY 195 YRAVRHSGGP-----VAWEHTYPEHTHLN---ETIIDEFATGPIG-----L 236
DB 720 YKVEVISMAGPLTYAANVASMTYPLIPNELVSMQAGSAVNLAMPSPGLGQACHAOL 779
QY 237 NEKLHSRSISRMYWRNFTPEFRAROKAISLRGASG 274
DB 780 SDAGHLS-----WE-----QPLKIGELFMLNDLTPG 806

RESULT 14

US-09-413-814-42
; Sequence 42, Application US/09413814
; Patent No. 6225064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bloecker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hoelle, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

CURRENT FILING DATE: 1999-10-07

EARLIER APPLICATION NUMBER: DE 198 46 493.2

EARLIER FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 107

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 42

LENGTH: 2539

TYPE: PRT

ORGANISM: Sorangium cellulosum

Query Match 4.8%; Score 77; DB 4; Length 2539;
Best Local Similarity 22.2%; Pred. No. 78;
Matches 42; Conservative 20; Mismatches 57; Indels 70; Gaps 8;

QY 86 HIYVYALKNPVCLTDARRRPINLLARVEOGLCDVGLGDASYGHRITKNPLSTAHATLMP 145
DB 1586 HLVOA-----MAEAGSYSHIRLLI-VSCGVYDVYGAE-----PQPARATMIGP 1628
QY 146 ADAL---YELRALHT-----LDEIHALPEAGNRRRVNYSYGRNVTL-----F 187
DB 1629 ALCTIPQYPHLETSHVDLVGVHADELHAROLDLSLAECISATVAERQALRGRRHMLDY 1688

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:40 ; Search time 20.69 Seconds
(without alignments)
100.347 Million cell updates/sec

Title: US-09-720-583a-3

Perfect score: 420
Sequence: 1 MTTRERLPRNGYSIAAAAKK.....EVGVSVGVVYALNKNRTDA 85

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTus.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.5	16.1	375	2 US-08-446-875-8	Sequence 8, Appli
2	67.5	16.1	375	2 US-08-102-385G-8	Sequence 8, Appli
3	62	14.8	328	2 US-08-956-012-1	Sequence 1, Appli
4	61.5	14.6	130	2 US-08-494-907-20	Sequence 20, Appli
5	61.5	14.6	130	5 PCT-US96-10986-20	Sequence 20, Appli
6	61.5	14.6	202	2 US-08-494-907-16	Sequence 16, Appli
7	61.5	14.6	202	5 PCT-US96-10986-16	Sequence 16, Appli
8	61	14.5	100	2 US-08-160-524A-12	Sequence 12, Appli
9	60.5	14.4	114	2 US-08-473-020A-26	Sequence 26, Appli
10	60	14.3	872	1 US-08-491-357-3	Sequence 3, Appli
11	60	14.3	872	3 US-08-968-633-3	Sequence 3, Appli
12	60	14.3	872	3 US-09-196-466-3	Sequence 3, Appli
13	60	14.3	872	5 PCT-US96-10823-3	Sequence 3, Appli
14	59	14.0	450	3 US-09-071-434-1	Sequence 1, Appli
15	58	13.8	977	3 US-08-335-844A-22	Sequence 22, Appli
16	57.5	13.7	553	4 US-09-413-814-3	Sequence 3, Appli
17	57	13.6	261	1 US-08-245-688-2	Sequence 2, Appli
18	57	13.6	261	1 US-08-245-688-6	Sequence 6, Appli
19	57	13.6	261	1 US-08-245-688-12	Sequence 12, Appli
20	56.5	13.5	261	1 US-08-245-688-4	Sequence 4, Appli
21	56.5	13.5	261	1 US-08-245-688-8	Sequence 8, Appli
22	56.5	13.5	261	1 US-08-245-688-10	Sequence 10, Appli
23	56.5	13.5	288	1 US-08-375-709-9	Sequence 9, Appli
24	56.5	13.5	288	1 US-08-752-929-9	Sequence 9, Appli
25	56.5	13.5	350	4 US-09-090-793-6	Sequence 6, Appli
26	56.5	13.5	350	4 US-09-655-270A-17	Sequence 17, Appli
27	56.5	13.5	350	4 US-09-651-941-21	Sequence 21, Appli

28	56.5	13.5	781	2 US-08-675-631-3	Sequence 3, Appli
29	56	13.3	80	4 US-08-737-226-5	Sequence 5, Appli
30	56	13.3	245	4 US-09-321-276-2	Sequence 2, Appli
31	56	13.3	245	4 US-08-916-481-5	Sequence 5, Appli
32	56	13.3	410	3 US-08-948-997-2	Sequence 2, Appli
33	56	13.3	410	4 US-09-348-817A-2	Sequence 2, Appli
34	55.5	13.2	444	4 US-08-822-774-52	Sequence 52, Appli
35	55.5	13.2	444	4 US-09-632-711-52	Sequence 52, Appli
36	55	13.1	244	2 US-08-553-633A-5	Sequence 3, Appli
37	55	13.1	298	2 US-09-006-535-3	Sequence 3, Appli
38	55	13.1	1201	4 US-09-098-901-2	Sequence 2, Appli
39	54.5	13.0	573	1 US-08-200-512-2	Sequence 2, Appli
40	54.5	13.0	2756	1 US-08-375-709-11	Sequence 11, Appli
41	54.5	13.0	2756	1 US-08-752-929-11	Sequence 11, Appli
42	54.5	13.0	2756	4 US-09-090-793-7	Sequence 7, Appli
43	53.5	12.7	224	1 US-08-608-241-4	Sequence 4, Appli
44	53.5	12.7	224	1 US-08-922-182-4	Sequence 4, Appli
45	53.5	12.7	224	2 US-08-919-953-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-446-875-8
Sequence 8, Application US/08446875
Patent No. 5858751
GENERAL INFORMATION:
APPLICANT: Paulson, James C.
APPLICANT: Wen, Xiaohong
APPLICANT: Livingston, Brian Duane
APPLICANT: Gillespie, William
APPLICANT: Kelm, Sorge
APPLICANT: Burlingame, Alma L.
APPLICANT: Medzihradszky, Katalin
TITLE OF INVENTION: Compositions and Methods for the
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESS: Poms, Smith, Lande & Rose
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,875
FILING DATE: July 12, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/102,385
FILING DATE: August 4, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 111-197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-875-8

Query Match 16.1%; Score 67.5; DB 2; Length 375;
Best Local Similarity 25.3%; Pred. No. 0.75;
Matches 24; Conservative 13; Mismatches 31; Indels 27; Gaps 4;

QY 2 TTRERLRNGYSIAAAKLGVS-----ESTVLRWTSSEPREEVAVARHARIRELRSE 56
DB 52 SNRAEVINGSSLPAAVANSNESLKHSTIPASSKW-----RHNOTLSLRIR 97

QY 57 GOSMRAIAAEVGSV--GT-----VHYALNKNRT 83
DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNNRST 132

RESULT 2
US-08-102-385G-8
; Sequence 8, Application US/08102385G
; Patent No. 5962294
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane
; APPLICANT: Gillespie, William
; APPLICANT: Kelm, Jorge
; APPLICANT: Burlingame, Alma L.
; APPLICANT: Medzhitskyy, Katalin
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Openheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, 38th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,385G
; FILING DATE: 04-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925369
; FILING DATE: 04-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oidenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 97-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310)788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-385G-8

Query Match 16.1%; Score 67.5; DB 2; Length 375;
Best Local Similarity 25.3%; Pred. No. 0.75;
Matches 24; Conservative 13; Mismatches 31; Indels 27; Gaps 4;

QY 2 TTRERLRNGYSIAAAKLGVS-----ESTVLRWTSSEPREEVAVARHARIRELRSE 56
DB 52 SNRAEVINGSSLPAAVANSNESLKHSTIPASSKW-----RHNOTLSLRIR 97

QY 57 GOSMRAIAAEVGSV--GT-----VHYALNKNRT 83
DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNNRST 132

DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNNRST 132

RESULT 3
US-08-956-012-1
; Sequence 1, Application US/08956012
; Patent No. 5972660
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN HYDROXYPIRUVATE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,012
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0410 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-845-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-956-012-1

Query Match 14.8%; Score 62; DB 2; Length 328;
Best Local Similarity 26.4%; Pred. No. 3.3;
Matches 23; Conservative 16; Mismatches 28; Indels 20; Gaps 4;

QY 6 RLPRNGYSIAAAKKGVSSTVLRWTSSE---PREEVAVARHARI-----REL 53
DB 14 RPAEGRVALARA-----ADCEVEQWSDDEIPAKELERGVAGAHGILCLSDHDKRIL 68

QY 54 RSEGSRAIAAEVGSVGTVHYALNK 80
DB 69 DAAGANLKVIST---MSVGIDHLALDE 92

RESULT 4
US-08-494-907-20
; Sequence 20, Application US/08494907
; Patent No. 5955298
; GENERAL INFORMATION:
; APPLICANT: Thomasow, Linda S
; APPLICANT: Bangera, Mahalaxmi
; APPLICANT: Weller, David M
; APPLICANT: Cook, R. James
; TITLE OF INVENTION: Sequences for Production of
; TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
; NUMBER OF SEQUENCES: 20


```

Db      36  SGLSIEVAARRGASCKPTIYRWGNK-----AALIAEYESSESEQIRKEPDKGSFKE 87
              :||| ||: | |: | ||           || | : | ||
Oy      57  -----GQSMRAIAAEVGSVGIV 74
              ||: | : | | : | |
Db      88  NLNELLNLIMKYWRETCGEAFRCVIAEAQLDPSTL 123
                                         GOSMRAIAAEVGSVGIV 74
                                         ||: | : | | : | |
Oy      57  -----GOSMRAIAAEVGSVGIV 74
                                         ||: | : | | : | |
Db      88  NLNELLNLIMKYWRETCGEAFRCVIAEAQLDPSTL 123
                                         ||: | : | | : | |

RESULT    7
PCT-US96-10986-16
; Sequence 16, Application PC/TUS9610986
; GENERAL INFORMATION:
; TITLE OF INVENTION: Sequences for Production of
; TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Stephan A. Pendorf, DOMINIK & STEIN
; STREET: 600 N. West Shore Boulevard, Suite 1000
; CITY: Tampa
; STATE: FL
; COUNTRY: USA
; ZIP: 33609
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10986
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pendorf, Stephan A.
; REGISTRATION NUMBER: 32665
; REFERENCE/DOCKET NUMBER: A700.320
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (813) 289-2966
; TELEFAX: (813)289-2967
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10986-16

Query Match          14.6%; Score 61.5; DB 5; Length 202;
Best Local Similarity 21.9%; Pred.No.2;
Matches   21; Conservative   11; Mismatches   25; Indels   39; Gaps     2

Oy      10  NGYSTAAAKKIAGVESITYKRWTSEPREFVARVAARHARELNSE-----56
             | | | | : | : | | | | | | | | | | | | | | | |
Db      36  SGLSIEVAARRGASKPTIYRWGNK-----AALIAEYESSESEQIRKEPDKGSFKE 87
             | | | | : | : | | | | | | | | | | | | | | | |

Oy      57  -----GOSMRAIAAEVGSVGIV 74
             ||: | : | | : | |
Db      88  NLNELLNLIMKYWRETCGEAFRCVIAEAQLDPSTL 123
             ||: | : | | : | |

RESULT    8
US-08-160-524A-12
; Sequence 12, Application US/08160524A
; Patent No. 5851761
; GENERAL INFORMATION:
; APPLICANT: McAdam, Ruth Anne
; APPLICANT: Dale, Jeremy W.
; APPLICANT: Zainuddin, Zaiarul Fadziruddin B.
; APPLICANT: Catly, David
; TITLE OF INVENTION: PROBES, KITS AND METHODS FOR THE
; TITLE OF INVENTION: DETECTION AND DIFFERENTIATION OF MYCOBACTERIA
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:

```

```

1 ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert,
2 ADDRESSEE: Attn: Walter H. Dreger
3 STREET: 4 Embarcadero Center, Suite 3400
4 CITY: San Francisco
5 STATE: California
6 COUNTRY: United States
7 ZIP: 94111-4187
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 OPERATING SYSTEM: IBM PC compatible
12 SOFTWARE: Patentln Release #1.0, Version #1.25
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/160,524A
15 FILING DATE: 01-DEC-1993
16 CLASSIFICATION: 435
17
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 07/752,661
20 FILING DATE: 18-OCT-1991
21
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: GB 8903968.9
24 FILING DATE: 22-FEB-1989
25
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: GB 9000411.0
28 FILING DATE: 09-JAN-1990
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: PCT/GB90/00276
32 FILING DATE: 22-FEB-1990
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Dreger, Walter H.
36 REGISTRATION NUMBER: 24,190
37 REFERENCE/DOCKET NUMBER: A-55387-1/WHD
38
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: (415) 781-1989
41 TELEFAX: (415) 398-3249
42
43 TELEX: 910 277299
44
45 INFORMATION FOR SEQ ID NO: 12:
46
47 SEQUENCE CHARACTERISTICS:
48 LENGTH: 100 amino acids
49 TYPE: amino acid
50 TOPOLOGY: unknown
51
52 US-08-160-524A-12
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
94
```


HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-968-633-3

Query Match 14.3%; Score 60; DB 3; Length 872;
Best Local Similarity 34.5%; Pred. No. 23;
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVARHARRIRIELRSEGSQMRALAEVGSVGVYHALNK 80
| | | | | : | | | | : | | | | : | | | | : | | | | :
DB 493 RSTSEQPEPPVDLKAANAAGVAVHGVHLELFARS--AVSSATHSDRTLHAKLSR 545

RESULT 12
US-09-196-466-3
; Sequence 3, Application US/09196466
; Patent No. 6124434
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, Joanne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,466
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-09-196-466-3

Query Match 14.3%; Score 60; DB 3; Length 872;
Best Local Similarity 34.5%; Pred. No. 23;
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVARHARRIRIELRSEGSQMRALAEVGSVGVYHALNK 80
| | | | | : | | | | : | | | | : | | | | : | | | | :
DB 493 RSTSEQPEPPVDLKAANAAGVAVHGVHLELFARS--AVSSATHSDRTLHAKLSR 545

RESULT 13

PCT-US96-10823-3
; Sequence 3, Application PC/TUS9610823
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, Joanne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10823
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US96-10823-3

Query Match 14.3%; Score 60; DB 5; Length 872;
Best Local Similarity 34.5%; Pred. No. 23;
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVARHARRIRIELRSEGSQMRALAEVGSVGVYHALNK 80
| | | | | : | | | | : | | | | : | | | | : | | | | :
DB 493 RSTSEQPEPPVDLKAANAAGVAVHGVHLELFARS--AVSSATHSDRTLHAKLSR 545

RESULT 14
US-09-071-434-1
; Sequence 1, Application US/09071434
; Patent No. 6080848
; GENERAL INFORMATION:
; APPLICANT: Sven, H
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN BRAIN-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,434
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0517 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT03
; CLONE: 662575
;
US-09-071-434-1

Query Match      14.0%; Score 59; DB 3; Length 450;
Best Local Similarity 31.9%; Pred. No. 13;
Matches 22; Conservative 11; Mismatches 30; Indels 6; Gaps 3;

QY 14 IAAAKKLGV---SESVYKRWTSRPREFVARVARHARRIRELRSEGSMAITAEGVSGVGVHYALNK 70
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 140 VLETVRLEGAIVTLETLLEVELEPRTELVA--AARGAR-RQAEAAAOLOGLAFWOGVP 196
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 71 VGTVHYVALN 79
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 197 LSPLOVAEN 205

RESULT 15
US-08-335-844A-22
; Sequence 22, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MUNN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,844A
; FILING DATE: 09-JAN-1995
```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9209936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 977 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-335-844A-22

Query Match      13.8%; Score 58; DB 3; Length 977;
Best Local Similarity 28.1%; Pred. No. 50;
Matches 16; Conservative 14; Mismatches 19; Indels 8; Gaps 2;

QY 24 SESYVKKRWTSRPREFVARVARHARRIRELRSEGSMAITAEGVSGVGVHYALNK 80
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 3 AEESEQEOTQOPKRNVTYLRLP---IKSL---FALLVAAAVGLSIGLYTFTRK 51

Search completed: October 4, 2002, 15:33:41
Job time: 33 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:36:07 ; Search time 46.38 Seconds

(without alignments)
317.045 Million cell updates/sec

Title: US-09-720-583A-3

Perfect score: 420
Sequence: 1 MTRERLRPRNGYSIAAAKK.....EVGVSQVTVHYALNKNRTDA 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPRREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	420	100.0	85	2	09F6K0
2	215	51.2	87	2	046076
3	162	38.6	67	2	09F4V5
4	140	33.3	93	2	09LBJ5
5	121	28.8	94	2	09F5N5
6	110	26.2	93	2	030682
7	96	22.9	114	2	09L456
8	82.5	19.6	105	2	045287
9	82	19.5	119	2	052144
10	79	18.8	87	2	093Q22
11	78	18.6	132	2	09EX03
12	78	18.6	150	17	097ZM9
13	78	18.6	318	2	086758
14	78	18.6	318	2	09RJU2
15	78	18.6	338	2	09X467
16	78	18.6	339	2	09K393

17	78	18.6	339	2	09L0Y8	0910Y8 streptomyce
18	78	18.6	339	2	09L0E4	0910E4 streptomyce
19	78	18.6	941	2	09Z1W5	09Z1W5 streptomyce
20	76	18.1	328	2	P94642	P94642 corynebacte
21	76	18.1	329	2	030997	030997 corynebacte
22	75.5	18.0	314	2	085929	085929 sphingomona
23	75	17.9	293	16	09ZRW5	09ZRW5 rhizobium m
24	74.5	17.7	195	2	09ZHM7	09ZHM7 brucella me
25	74.5	17.7	674	2	047763	047763 enterococcu
26	74.5	17.7	679	2	09S6C2	09S6C2 enterococcu
27	74	17.6	139	17	09Y132	09Y132 pyrococcus
28	74	17.6	310	16	09Z1V4	09Z1V4 rhizobium m
29	73.5	17.5	128	2	085930	085930 sphingomona
30	73	17.4	102	16	098B11	098B11 rhizobium 1
31	73	17.4	122	2	088091	088091 escherichia
32	72.5	17.3	184	16	09ZS36	09ZS36 rhizobium m
33	72.5	17.3	315	2	0989L5	0989L5 rhizobium m
34	72.5	17.3	315	2	09X8B3	09X8B3 rhizobium m
35	72.5	17.3	643	17	090Y87	090Y87 pyrococcus
36	72.5	17.3	673	2	047751	047751 enterococcu
37	72.5	17.3	673	2	047783	047783 enterococcu
38	72.5	17.3	678	2	059438	059438 enterococcu
39	72.5	17.3	678	2	047759	047759 enterococcu
40	72.5	17.3	678	2	047801	047801 enterococcu
41	72.5	17.3	678	2	093N65	093N65 enterococcu
42	72.5	17.3	678	2	093NP3	093NP3 enterococcu
43	72	17.1	88	2	068709	068709 yersinia pe
44	72	17.1	113	2	053005	053005 escherichia
45	72	17.1	125	16	099Q10	099Q10 caulobacter

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	85 AA.
ID	09F6K0			
AC	09F6K0			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)			
DE	REP2.			
OS	Propionibacterium freudenreichii.			
OG	Plasmid p545.			
OC	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Propionibacterineae; Propionibacteriaceae;			
OC	Propionibacterium.			
OX	NCBI_TaxID=1744;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=LMG16545;			
RX	MEDLINE=21091935; PubMed=11157209;			
RA	Jore J.P.M., van Luijk N., Luiten R.G.M., van der Werf M.J.,			
RA	Pouwels P.H.;			
RT	"Efficient transformation system for Propionibacterium freudenreichii			
RT	based on a novel vector";			
RL	Appl. Environ. Microbiol. 67:499-503(2001).			
DR	EMBL: AF291751; AAG25295.1; -.			
KW	Plasmid.			
SQ	SEQUENCE 85 AA; 9425 MM; 30FBBA897F0CC6F0 CRC64;			
Query Match	100.0%;	Score 420;	DB 2;	Length 85;
Best Local Similarity	100.0%;	Pred. No. 2.2e-35;		
Matches	85;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
QY	1	MTRERLRPRNGYSIAAAKLGVSSTVKKWTSPPREFVARYAARHARIRREISQSGM 60		
DB	1	MTRERLRPRNGYSIAAAKLGVSSTVKKWTSPPREFVARYAARHARIRREISQSGM 60		
QY	61	RAIAAEVGSVGTVHYALNKNRTDA 85		
DB	61	RAIAAEVGSVGTVHYALNKNRTDA 85		

RESULT	2			
046076				
ID	046076	PRELIMINARY;	PRT;	87 AA.
AC	046076			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHEETICAL 9.8 KDA PROTEIN.			
OS	Corynebacterium glutamicum (Brevibacterium flavum).			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Corynebacteriaceae;			
OC	Corynebacterium			
OX	NCBI_Taxid-1718;			
KN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-1014;			
RA	Shi S.L., Wang Z.X., Deng Y., Zhu H., He B., Zheng Z.X.;			
RT	"Complete Nucleotide Sequence of a Plasmid pXZ10142 from			
RT	Corynebacterium glutamicum."			
RL	Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: X72691; CAAS1239.1; -.			
KM	Hypothetical protein.			
SQ	SEQUENCE 87 AA; 9769 MW; 86DA0A01601185747 CRC64;			

[illegible]

RESULT	3			
09F4V5				
ID	09F4V5	PRELIMINARY:	PRT:	67 AA.
AC	09F4V5.			
DT	01-MAR-2001 (TREMBlrel. 16, Created)			
DT	01-MAR-2001 (TREMBlrel. 16, last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, last annotation update)			
DE	REPB.			
GN	REPB.			
OS	Brevibacterium linens.			
OG	Plasmid f1M.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Micrococciaceae; Brevibacteriaceae; Brevibacterium.			
OK	NCBI_TaxID=1703;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Moore M.W., Bowling D., Glenn D.;			
RT	"Nucleotide sequence of a plasmid from Brevibacterium linens:			
RT	Construction of a cloning vector.";			
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBDJ databases.			
DR	EMBL, AY004211; AAF89086.1; --			
KW	Plasmid.			
50	SEQUENCE	67 AA;	7850 MW;	CF7DDE3EA6EA4170 CRC64;

```

Query Match      38.6%  Score 162; DB 2; Length 67;
Best Local Similarity 54.2%  Pred. No. 2,3e-09;
Matches 32; Conservative 7; Mismatches 20; Indels 0; Gaps 0;

OY 1 MTTTRRLRNGYSIAAAAKKLCVSESTYKRWTSERPEEFVAVAAARHARIRLRSGQS 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 MTAERHPHPNGSVLELAKVGSSEKSYLRWTSERPEVLLSAQGRVRAIRRLRGTGLS 61
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT	4		
09LBJ5			
ID	09LBJ5	PRELIMINARY;	PRT; 93 AA.
AC	09LBJ5		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	DNA-BINDING REPLICATION PROTEIN.		
GN	REP.		
OS	Rhodococcus rhodochrous.		
OC	plasmid pRC4.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;		
OC	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.		
OX	NCBI_taxid:1829;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-IF03338.		
RX	MEDLINE=21196497; PubMed=11302154;		
RA	Hirasawa K., Ishii Y., Kobayashi M., Koizumi K., Maruhashi K.;		
RT	"Improvement of Deutilization Activity in Rhodococcus erythropolis		
RT	Ka2-5-1 by Genetic Engineering."		
RL	Biosci. Biotechnol. Biochem. 65:239-246(2001).		
DR	EMBL: AB040101; BAA92819.1; -		
DR	InterPro: IPR000792; HTH_LuxR.		
KW	DNA-binding; Plasmid.		
SO	SEQUENCE 93 AA; 10124 MW; D328D70CFA51A21A CRC64;		

Query Match	33.3%	Score 140;	DB 2;	Length 93;
Best Local Similarity	43.1%;	Pred. 0.57e-07;		
Matches	31;	Conservative 14;	Mismatches 23;	Indels 4;
				Gaps 1;
Oy	18	AKKLGVSESTVWRMTSEPREEFVAVARHARIRLREJSEGSOMRAIAEVEGVSTV---74		
Db	17	AERIASPPTVARRIIAEPRASTEAPAAERKOVLELRASGMKRLREIAAEVGSVGVGTI76		
Oy	75	-HYALNKRTDA85		
Db	77	LHHAARKTEOSKA88		

RESULT	5			
Q9F5N5				
ID	Q9F5N5	PRELIMINARY;	PTM;	94 AA.
AC	Q9F5N5;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation updates)			
DE	PUTATIVE DNA-BINDING PROTEIN.			
OS	Rhodococcus erythropolis.			
OG	Plasmid pN30.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.			
OX	NCBI_TaxID=1693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ryabchenko L.E., Novikov A.D., Golyshtin P.N., Yansenko A.S.,			
RT	"Rhodococcus erythropolis plasmid pN30 putative replicase (ORF1), and			
RT	putative DNA-binding replication protein (ORF2)."			
RL	Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF312210; AAC29855.1; -			
KM	DNA-binding; Plasmid.			
SO	SEQUENCE 94 AA; 10409 MW; B3452537C39166E7 CRC64;			

Query Match	28.8%	Score 121	DB 2	Length 94			
Similarity	37.8%	Pred. NO. 4.9e-05					
Best Local							
Matches	28	Conservative 14	Mismatches 33	Indels 0			
Gaps	0						
Qy	9	RNGCYIAAAAKKLGVS	ESTVKRWITSEPR	EYFVARYAARHARIR	ELRSLSECSQSR	RAITAAEYVG	68
		: :	: :	: :	: :	: :	
Db	9	RKKMTAAAAAEFFG	ASTRIQRLFAEP	RNDYDYGRAK	KARDKVELR	KQGLKRYTAE	68
		:	:	:	:	:	


```

QY 69 VSVGVTHYALKNNR 82
      :| | | | |
DB 69 LSTGIVGRLLHDAK 82

RESULT 6
ID 030682 PRELIMINARY; PRT; 93 AA.
AC 030682;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE DNA-BINDING REPLICATION PROTEIN.
GN REPB.
OS Rhodococcus erythropolis.
OC Plasmid pFAJ2600.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N186/21;
RX MEDLINE=98015402; PubMed=9353918;
RA De Moe R., Nagy I., De Schrijver A., Pattanapitpaisal P.,
RA Schoofs G., Vanderleyden J.;
RT "Structural analysis of the 6-kb cryptic plasmid pFAJ2600 from
RT Rhodococcus erythropolis N186/21 and construction of Escherichia coli-
RT Rhodococcus shuttle vectors.";
RL Microbiology 143:3137-3147(1997).
DR EMBL; AF013088; AAC45811.1; -.
KW DNA-binding; Plasmid.
SQ SEQUENCE 93 AA; 10672 MW; 7AFAB05CE8EE5ED CRC64;

Query Match 26.28; Score 110; DB 2; Length 93;
Best Local Similarity 35.78; Pred. No. 0.00063;
Matches 25; Conservative 12; Mismatches 33; Indels 0; Gaps 0;

QY 5 ERPLRNGYSIAAAKKLGVESTVKRWTSPEEFVAVARHARIRBLR-SEGQSMRA 64
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4 ENOTRRKRTAKEMAEKRGVSERTIRNIVASRDSYQAAAEERRDTAVKLEQGMKRYEIA 63
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 65 AEVGVSVGVTV 74
      :| | | | |
DB 64 EEMEISTGAV 73

RESULT 7
ID 091456 PRELIMINARY; PRT; 114 AA.
AC 091456;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHETICAL 13.0 KDA PROTEIN (REPLICATION PROTEIN).
GN REPB.
OS Propionibacterium jensenii, and
OS Propionibacterium acidipropionici.
OC Plasmid pIME106, and Plasmid pRCO1.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Propionibacterineae; Propionibacteriaceae;
OC Propionibacterium.
OX NCBI_TaxID=1749, 1748;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=P.jensenii; STRAIN=DF1; PLASMID=pIME106;
RA Stierli M.P., Welle U., Teuber M.;
RT "Molecular analysis of plasmid pIME106, a plasmid isolated from
RT Propionibacterium jensenii.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
[2]
SQ SEQUENCE FROM N.A.

```

```

RC SPECIES=E.acidipropionici; STRAIN=E214; PLASMID=pRCO1;
RX MEDLINE=20509786; PubMed=11055911;
RA Kiatpapap P., Hashimoto Y., Nakamura H., Piao Y.Z., Ono H.,
RA Yamashita M., Murooka Y.;
RT "Characterization of pRCO1, a plasmid from Propionibacterium
RT acidipropionici, and its use for development of a host vector system
RT in Propionibacterium.";
RL Appl. Environ. Microbiol. 66:4688-4695(2000).
DR EMBL; AJ250233; CAB88397.1; -.
DR EMBL; AB007909; BAB17920.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 114 AA; 13048 MW; 6517E7A9C3F304B2 CRC64;

Query Match 22.98; Score 96; DB 2; Length 114;
Best Local Similarity 29.38; Pred. No. 0.021;
Matches 24; Conservative 17; Mismatches 39; Indels 2; Gaps 2;

QY 4 ERPLRNGYSIAAAKKLGVESTVKRWTSPEEFVAVARHARIRBLR-SEGQSMRA 62
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5 QHRVRRRG-TAREAAERVGASIRTAQWTSIPREEMWITQKAVEDEIRAYKDEGHWTGE 63
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 63 IAAEVGVSVGVTHYALKNNRTD 84
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 64 TSHHFGIAKTYAQBERRARRRE 85

RESULT 8
ID 045287 PRELIMINARY; PRT; 105 AA.
AC 045287;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ORF1 PROTEIN.
GN ORF1.
OS Bifidobacterium longum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=1679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B2577;
RX MEDLINE=96341130; PubMed=8761732;
RA Rossi M., Brigidi P., Gonzalez Vara Y Rodriguez A., Matteuzzi D.;
RT "Characterization of the plasmid pmb1 from Bifidobacterium longum and
RT its use for shuttle vector construction.";
RL Res. Microbiol. 147:133-143(1996).
DR EMBL; X84655; CAA59149.1; -.
SQ SEQUENCE 105 AA; 12317 MW; 295DD15830188041 CRC64;

Query Match 19.68; Score 82.5; DB 2; Length 105;
Best Local Similarity 28.78; Pred. No. 0.44;
Matches 25; Conservative 13; Mismatches 42; Indels 7; Gaps 3;

QY 2 TTRERLPRNGYSIAAAKKLGVESTVKRWTSPEEFVAVARHARIRBLR-SEGQSM 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5 TLKKRPVSARELAE---YGVSTRITOSVYAKMRREDWIDQAAAMREAVRYHDEGHWTW 61
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 RAIAEVGVSVGVTH---YALNNKRTD 84
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 PQTAEHFNMGSQAVRQRCYRARKERED 88

RESULT 9
ID 052144 PRELIMINARY; PRT; 119 AA.
AC 052144;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL 13.7 KDA PROTEIN.

```



```
Db 11 KMYEGLSTRELANOLGSLYSKYRRLIKAKVNFCKVPNDKIÖKIEMKOCYSANRI 70
OY 64 AAEVGSVGTV-----HYALNKNR 82
      : : : : :
Db 71 SRELINENTVTLRLKKYNLNGKKR 94

RESULT 13
ID 086758 PRELIMINARY; PRT; 318 AA.
AC 086758;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE TRANSPOSASE.
GN SC6A9.26 OR SCJ11.20.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL109949; CAB52905.1; -.
DR EMBL; AL109949; CAB52905.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; Rve; 1.
SO SEQUENCE 318 AA; 36530 MW; 61446B9CA1DBEE9 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 318;
Best Local Similarity 27.4%; Pred. No. 4.4;
Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

OY 2 TTRELRPR-----NGYSIAAAKRLGVSESTVKRW-----SEPR-- 36
      | | | | : : : : :
Db 11 TGRRLARCVEDGWPVRAAERFQVSHTTASRNARKYRQLGYGMSDRSRPHHQPRRT 70
OY 37 -----EEFVAARVAARHARIRELRSGOSMRALAAEVGSVGTVAHAL 78
      | | | | : | | | : | | | : | | |
Db 71 AAVEEHVLRRLREH-RIGPLR-----LAVRCGIAASTAHRIL 107

RESULT 14
OY 14 PRELIMINARY; PRT; 318 AA.
ID 09RJU2
AC 09RJU2;
```

```
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE TRANSPOSASE (PUTATIVE INSERTION ELEMENT TRANSPOSASE).
GN SCF41.27 OR SC10B8A.13.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117387; CAB5730.1; -.
DR EMBL; AL391454; CAB5730.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; Rve; 1.
SO SEQUENCE 318 AA; 36544 MW; 65546B9CA1DBEE9 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 318;
Best Local Similarity 27.4%; Pred. No. 4.4;
Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

OY 2 TTRELRPR-----NGYSIAAAKRLGVSESTVKRW-----SEPR-- 36
      | | | | : : : : :
Db 11 TGRRLARCVEDGWPVRAAERFQVSHTTASRNARKYRQLGYGMSDRSRPHHQPRRT 70
OY 37 -----EEFVAARVAARHARIRELRSGOSMRALAAEVGSVGTVAHAL 78
      | | | | : | | | : | | | : | | |
Db 71 AAVEEHVLRRLREH-RIGPLR-----LAVRCGIAASTAHRIL 107

RESULT 15
OY 15 PRELIMINARY; PRT; 338 AA.
ID 09X467
AC 09X467;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE TRANSPOSASE.
GN TNP4.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN-A3(2);
 RX MEDLINE-99250253; PubMed-10231572;
 RA Chung H.J., Kim E.J., Suh B., Choi
 RT "Duplicate genes for Fe-containing superoxide dismutase in
 ST Streptomyces coelicolor A3(2).";
 RL Gene 231,87-93(1999);
 DR EMBL; AF099014; AAD33127.1; -;
 DR InterPro; IPR01584; Rve.
 DR Pfam; PF00665; rve; 1.
 SO SEQUENCE 338 AA; 38807 MW; E92CD4811DOC615 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 338;
 Best Local Similarity 27.4%; Pred. No. 4.7;
 Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

QY 2 TTRERLP-NGYSIAAAKKLGSESTVKRW-SEPR-- 36
 Db 32 TGRRLACVYEDGFPVRAAERFQVSHHTASRWARRYQLGVTGMSDRSSRPHHQPRRT 91
 QY 37 ----EEFVARYAARHARIRELRSEQSMRAIAAEVGVSGTVHYAL 78
 Db 92 AAAYVEHYLRLEH-RIGPLR-----LAVRCGIAASTAHRIL 128

Search completed: October 4, 2002, 15:36:08
 Job time: 180 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:34:40 ; Search time 55.08 Seconds
(without alignments)
171.410 Million cell updates/sec

Title: US-09-720-583A-3

Perfect score: 420
Sequence: 1 MTRERLRPRNGYSIAAAAKR.....EVCYSGTVHVAIKNRDTA 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_032802:*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	420	100.0	85	21	AAV44637
2	82	19.5	314	18	AAW14834
3	79	18.8	314	13	AAK29623
4	79	18.8	314	13	AAK37873
5	79	18.8	314	13	AAK34544
6	76	18.1	314	13	AAK20991
7	69.5	16.5	316	21	AAAB14150
8	67	16.0	98	22	AAAG91658
9	67	16.0	279	22	AAU44478
10	67	16.0	927	22	AAE10133
11	66	15.7	214	22	AAU59374

12	65.5	15.6	132	22	AAU58347
13	65	15.5	168	22	ABB64106
14	64	15.2	162	22	AAU56724
15	64	15.2	743	22	ABB68417
16	64	15.2	1100	22	ABG24905
17	63.5	15.1	98	22	AAAG90746
18	63.5	15.1	98	22	AAAG90753
19	63.5	15.1	186	18	AAW30754
20	63.5	15.1	186	19	AAW50189
21	63.5	15.1	186	19	AAW44288
22	63.5	15.1	186	19	AAW44289
23	63.5	15.1	186	19	AAW44286
24	63.5	15.1	186	19	AAW44287
25	63.5	15.1	211	22	AAU52836
26	63.5	15.1	326	22	AAU34593
27	63.5	15.1	547	22	AAO13897
28	63.5	15.1	573	21	AAV82154
29	63.5	15.1	580	21	AAV82157
30	63.5	15.1	763	21	AAV82155
31	63.5	15.1	953	21	AAV82156
32	63	15.0	219	22	AAU63583
33	63	15.0	247	21	AAAG20359
34	63	15.0	279	21	AAAG20358
35	63	15.0	284	21	AAAG20357
36	63	15.0	2169	22	AAAG91701
37	62.5	14.9	507	22	AAAB96433
38	62	14.8	158	21	AAV81764
39	62	14.8	186	21	AAAG40791
40	62	14.8	204	21	AAAG40790
41	62	14.8	227	21	AAAG40789
42	62	14.8	328	20	AAAB43622
43	62	14.8	328	21	AAAB42600
44	62	14.8	356	22	AAU23556
45	62	14.8	362	22	AAU22915

ALIGNMENTS

RESULT 1	
ID	AAV44637 standard; Protein; 85 AA.
XX	
AC	AAV44637;
XX	
DT	07-APR-2000 (first entry)
XX	
DE	Proionibacterium LMG 16545 protein-2.
XX	
KW	Proionibacterium LMG 16545; vector; plasmid; antigen; vaccine; enzyme;
KW	nutritional factor; growth factor; clotting factor; antimicrobial; drug;
KW	hormone; vitamin B12; animal feed; lactic acid bacteria; foodstuff;
XX	
OS	Proionibacterium freudenreichii LMG 16545.
XX	
PN	WO967356-A2.
XX	
PD	29-DEC-1999.
XX	
PF	25-JUN-1999; 99MO-EP04416.
XX	
PR	25-JUN-1998; 98EP-030503.
XX	
PA	(KONN) GIST-BROCADES BV.
XX	
PI	Pouwels PH, Van Lujk N, Jore JPM, Luiten RGM,
XX	
DR	WPI; 2000-136977/12.
XX	
DR	N-PSDB; AA249691.
XX	
PT	Novel vectors containing Proionibacterium sequences, used to express homologous or heterologous proteins

XX	Claim 16; Page 26; 54pp: English.
PS	
XX	
CC	The present amino acid sequence is encoded by the endogenous plasmid
CC	Propiobacterium IMG 16545 ORF-2, derived from Propionibacterium
CC	freudenreichii deposited under CBS 101022 or CBS 101023. The vector has
CC	insertion sites for foreign DNA fragments and is capable of autonomous
CC	replication. It can be used safely as they remain extrachromosomal and
CC	are very stable. They can be used to express homologous or heterologous
CC	proteins, which may include antigens for use in vaccines, nutritional,
CC	growth and clotting factors, antimicrobials, vitamins (especially vitamin
CC	B12), enzymes, hormones and drugs. The Propionibacterium host cells are
CC	used in the production of animal feeds, manufacture of foodstuffs like,
CC	sausages and cheese, where they may be used instead of or in addition to
CC	lactic acid bacteria and in the manufacture of vitamin B12. The DNA
CC	sequence can also be used as a source of primers and probes.
XX	
SQ	Sequence 85 AA:
	Query Match 100.0%; Score 420; DB 21; Length 85;
	Best Local Similarity 100.0%; Pred. No. 3.2e-46;
	Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MTTTRRLPRNGYSIAAAKKLGSESYTKRWTSSEPFVAVAAARIRRLRSQSGM 60
Db	1 mtttrrlprnygsiaaakklgsestvkrtwtssepfvavaaarharirrlrsegsqm 60
OY	61 RAIAAEVGVSGVTGHVALNKNRTDA 85
Db	61 raiiaevgvsgvtghvalnknrtda 85
RESULT 2	
AAWI4834	
ID	AAWI4834 standard; Protein: 314 AA.
XX	
XX	AAWI4834:
XX	
XX	28-MAY-1997 (first entry)
XX	
DE	Protein encoded by M.rep ORE1 of plasmid pmV101.
XX	
XX	Plasmid pmV101; Mycobacteria; lipoprotein secretion signal sequence;
KW	antigen; Bacterial vaccine; Lyme disease; Borrelia burgdorferi;
KW	surface protein.
XX	
OS	Synthetic.
XX	
PN	US5583038-A.
XX	
PD	10-DEC-1996.
XX	
PF	21-OCT-1991; 91US-0780261.
XX	
PR	17-NOV-1992; 92US-0977630.
XX	
PR	21-OCT-1991; 91US-0780261.
XX	
PA	(MEDI-) MEDIMMUNE INC.
XX	
PI	Stover CK.
XX	
DR	WPI; 1997-042315/04.
DR	N-PSDB; AAT64413.
XX	
PT	Mycobacteria expressing secretion signal of lipoprotein and
PT	heterologous antigen, esp. outer surface protein A or B of Borrelia
PT	burgdorferi - are used in the form of a live bacterial vaccines
PT	against Lyme disease
XX	
PS	Example 1; Fig 5; 12pp: English.
XX	
CC	The sequences given in AAWI4834-36 are proteins which are encoded by

CC		the plasmid pMW101.	This plasmid was used in the construction of a
CC		series of vectors	to produce the recombinant Mycobacterium of
CC		The invention.	The recombinant mycobacteria of the invention are
CC		transformed with DNA	encoding a polypeptide which comprises a
CC		lipoprotein secretion signal sequence	and an antigen (Ag) heterologous
CC		to the mycobacteria.	The lipoprotein secretion signal causes the Ag
CC		to be produced as a lipoprotein.	The mycobacteria may be used in the
CC		form of a live bacterial vaccines against Lyme disease,	where the
CC		bacterial express a surface protein of Borrelia burgdorferi,	the
CC		causative agent of Lyme disease.	This sequence is printed in the
CC		specification in the C-terminal to N-terminal orientation as	determined from the corresponding DNA sequence.
SQ	Sequence	314 AA:	
Query Match		19.5%; Score 82; DB 18; Length 314;	
Best Local Similarity		36.1%;	Pred. No. 0.04;
Matches	26;	Conservative	8; Mismatches 34; Indels 4; Gaps 2;
OY	15	AAAKKLGVSSTVKRWTSSEPREFFVAVRANRHARIRELR-SEGOSMRAIAEVCVSQGT	73
Dd	223	aearltglsehvrvrlvagersewlaeaaarrrerirayhddeghspqtakhfglhldt	282
OY	74	VH---YALNKNR	82
Dd	283	vkrlygrarker	294
RESULT 3			
AAR29623	3	AAR29623 standard; Protein: 314 AA.	
XX	AC	AAR29623;	
XX	DT	04-JUN-1993 (first entry)	
DE	Tet e gene product from pMW101 - a deletion mutant of pTUB125.		
XX	Cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;		
KW	Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;		
XX	pertussis; malaria; influenza virus; CTL; herpes virus.		
OS	Mycobacterium.		
XX	WN9223376-A.	.	
PD	10-DEC-1992.		
PF	01-JUN-1992; 92WO-US04538.		
PR	06-JUN-1991; 91US-0711643.		
PA	(MED1-) MEDIMUNE INC.		
PI	Dela Cruz V, Stover CK;		
DR	WP1; 1992-433380/52.		
NR	N-PSDB; AAO31727.		
PT	Method of inducing cytotoxic T-lymphocyte response - esp		
PT	expression products of transformed Mycobacterium are useful as		
PM	vaccines against HIV, pertussis, malaria, influenza virus, herpes		
PM	virus, etc.		
XX			
PS	Example 1; Fig 5; 86pp; English.		
XX			
CC	This sequence is the tet e gene encoded by plasmid pMW101.		
CC			
SQ	Sequence	314 AA;	
Query Match		18.8%; Score 79; DB 13; Length 314;	


```

PR 16-JUL-1990; 90US-0553907.
XX
XX (YESH ) EINSTEIN A COLLEGE.
PA (UPLI-) UNIV OF PITTSBURGH.
XX
XX
XX Jacobs WR, Hatfull G;
PI
DR WPI: 1992-064943/08.
XX
XX DNA site-specific integration into Mycobacteria - useful as
PR adjuvant in vaccines and as therapeutic agent for malaria,
PR Influenza, herpes and human immunodeficiency virus
XX
XX Example 3; Fig 24; 82pp; English.
XX
XX Plasmid pYUB12 was constructed by ligating a 5kb Sau3 fragment from
CC pAL5000 (contg. a replication origin from M.fortitum) to BamHI-
CC digested pJ566 (contg. an E.coli ori and kanamycin-neomycin
CC resistance sequences). A 2586bp HpaI-EcoRV fragment from pYUB12
CC (comprising the minimum sequence necessary for plasmid replication
CC in BCG) was ligated to PvuII-digested pYUB8 to form pYUB53. (Plasmid
CC pYUB8 is a pBR322 deriv. which includes an E.coli replicon and a
CC kanamycin resistance gene). Twelve restriction sites were removed
CC by digesting pYUB53 with AatI, EcoRV and PstI. To eliminate 792
CC bases of the (inactivated) tet gene, the plasmid was digested with
CC NaeI and the 6407bp gel purified fragment was religated/
CC circularised. The resulting plasmid was designated pMV101.
CC The nucleotide sequence of plasmid pMV101 is printed in the
CC specification but the copy quality is too poor to allow the
CC sequence to be included on the GENESPO nucleotide database.
CC Three proteins are encoded by pMV101, i.e. the kanamycin resistance
CC protein and proteins "d" and "e" encoded by the complementary
CC strand. See also AAR20992 and AAR20993.
XX
XX Sequence 314 AA;
SQ

Query Match 18.1%; Score 76; DB 13; Length 314;
Best Local Similarity 34.7%; Pred. No. 0.24;
Matches 25; Conservative 8; Mismatches 35; Indels 4; Gaps 2;

OY 15 AAAAKKLGSESTVAKRMSTSEPREFVARVAAHRRRLRLR-SEGOSMRALAAVGVSGT 73
DB 223 aegallt1gsehvrlvqersewleagaartrrlrayhdqgshwptakfhghlhd 282
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 74 VH---YALNKNR 82
| | | | |
DB 283 VKrlgYrarker 294

RESULT 7
AAB14150
ID AAB14150 standard; Protein; 316 AA.
XX
XX AAB14150;
AC
XX 02-FEB-2001 (first entry)
DT
XX
XX Bordetella pertussis protein # 4.
DE
XX
XX Bordetella pertussis.
OS
XX
XX Bordetella pertussis.
PN
XX
XX WO200037493-A2.
XX
XX 29-JUN-2000.
PD
XX
XX 21-DEC-1999; 99WO-EPI0297.
PF
XX
XX 21-DEC-1998; 98GB-0028217.
PR
XX
XX (ULBR ) UNIV LIBRE BRUXELLES.
PA

```

XX	Bollen A, Fauconlier A, Godfroid E;
PI	WPI: 2000-452178/39.
DR	N-PSDB; AAA64888, AAA64890.
XX	
PI	Novel polypeptides derived from <i>Bordetella pertussis</i> , useful for
PT	treating and diagnosing <i>Bordetella</i> infection -
XX	
PS	Disclosure; Pages 160-161; 165pp; English.
XX	
CC	<i>Bordetella pertussis</i> possesses a type III secretion system. Type III
CC	secretion systems allow bacteria to target virulence factors directly at
CC	host cells. The present sequence is a protein of <i>B. pertussis</i> . No name
CC	has been specified for the present sequence. The coding sequence of the
CC	present sequence is homologous to housekeeping genes of other species,
CC	and so the present protein may have a role in cellular housekeeping. A
CC	pathogenicity island is a compact, distinct genetic unit carrying
CC	virulence genes. The coding sequence of the present sequence is located
CC	within a pathogenicity island (see AAA64890) which also carries a
CC	number of genes encoding proteins involved in the type III secretion
CC	system of <i>B. pertussis</i> . See AAA64849-A64884 and AAB1411-B14146 for
CC	details of the coding sequences and proteins identified in the
CC	pathogenicity island, of the present invention.
XX	
SQ	Sequence 316 AA;
	Query Match 16.5%; Score 69.5; DB 21; Length 316;
	Best Local Similarity 28.8%; Pred. No. 1.6;
	Matches 23; Conservative 10; Mismatches 32; Indels 15; Gaps 1;
OY	14 IAAAKKLGVSSESTVKRW-----TSEPREEFYARAAHARIRELRSEQ 58
	: : : : : : : : : :
Db	28 vpeaarayvtaptkyrkigrlagqgsladassrptvpratapakalaveirtrkl 87
OY	59 SMRAIAAEVGSVGTVHYAL 78
	: :
Db	88 tqariagalvvasctsvrll 107
RESULT	8
AAAG91658	
ID	AAAG91658 standard; Protein: 98 AA.
XX	
AC	AAAG91658;
XX	
XX	26-SEP-2001 (first entry)
DE	
XX	C glutamicum protein fragment SEQ ID NO: 5412.
XX	
KW	Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW	organic acid synthesis.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484.
XX	
PR	07-APR-2000; 2000JP-0159162.
XX	
PR	03-AUG-2000; 2000JP-0280988.
XX	
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
XX	WPI: 2001-376931/40.
DR	N-PSDB; AAH68877.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PS
XX
PS Claim 17; SEQ ID NO: 5412; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 98 AA;

Query Match 16.0%; Score 67; DB 22; Length 98;
Best Local Similarity 31.6%; Pred. No. 0.76;
Matches 18; Conservative 6; Mismatches 27; Indels 6; Gaps 1;

OY 11 GYSIAAAAKKLGVSESTVKKW-----TSEPRREFPARYAARARIRLRSEGSQNR 61
DB 23 gtsigkqkdiqjnmrlknwltkysanpshgntatalsearlrqlekenallr 79

RESULT 9
AAU44478
ID AAU44478 standard; Protein; 279 AA.
XX
AC AAU44478;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #5374.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Alonsoeuvre J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
DR N-PSDB; AAS59522.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX

PS Example 1; SEQ ID NO 5673; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 279 AA;

Query Match 16.0%; Score 67; DB 22; Length 279;
Best Local Similarity 36.5%; Pred. No. 2.9;
Matches 23; Conservative 11; Mismatches 25; Indels 4; Gaps 1;

OY 22 GVSSTVKKRWTSEPRREFPARYAARARIRLRSEGSQNRKAIAEV---GVSQGVTHYA 77
DB 10 gmatmvtgtsqjgrefvtrqlaargddlvvardtermaalkhadvearygsvetlaad 69

OY 78 LNK 80
DB 70 lsr 72

RESULT 10
AAE10133
ID AAE10133 standard; Protein; 927 AA.
XX
AC AAE10133;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysR3.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; transcriptional activator.
XX
OS Streptomyces noursei.
XX
FH Key
FH Domain
FT Location/Qualifiers
FT 26..47
FT /label= LZ
FT /note= "leucine zipper motif"
FT 548..568
FT /label= TMI
FT /note= "Transmembrane domain"
FT 583..610
FT /label= TM2
FT /note= "Transmembrane domain"
FT 684..912
FT /label= HTH
FT /note= "LuxR-type helix-turn-helix motif (DNA binding)"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.


```

PF 20-APR-2001; 2001WO-US12865.
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhattacharya A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI
DR WPI: 2001-616774/71.
DR N-PSDB: AAS59578.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID NO 17919; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 162 AA;
SQ
Query Match 15.2%; Score 64; DB 22; Length 162;
Best Local Similarity 30.4%; Pred. NO. 3.5;
Matches 17; Conservative 9; Mismatches 14; Indels 16; Gaps 2;
QY 12 YSIAAAKKLVSESTVAKRWTSERPEEFVAVRAAHARIRE-----LRSEGSMA 62
ID 11 : : | | | | : : : : : : : : : : : : : : : : : : : : : :
DB 31 hrvsdvaallgvsddtvttrwidq-----rhvtraakdttyrlvdgaslaa 75
XX
XX RESULT 15
XX ABB68417
XX AC ABB68417;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 32043.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX

```

XX	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
PA	(PEKE) PE CORP NY,
XX	
PI	Venter JC, Adams M, Li PMD, Myers EW;
XX	
DR	WPI, 2001-656860/75.
DR	N-PSDB; ABL12520.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Disclosure; SEQ ID NO 32043; 21np + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57731-ABB70272).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pt_sequences.
XX	
XX	Sequence 743 AA;
XX	

Query Match 15.2%; **Score** 64; **DB** 22; **length** 743;
Best Local Similarity 31.4%; **Pred.** NO. 25;
Matches 16; **Conservative** 15; **Mismatches** 14; **Indels** 6; **Gaps** 2;

QY	1 MTTRELP-----RNGISIAAAKRLGVSESTVKKM-TSEPPEEVARYYAA 45 :::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: 16 ltpcklqalgrlhndgeskaasvaaidlygwccknecllrmsrsga 66
----	---

Search completed: October 4, 2002, 15:34:41
Job time: 94 sec

RESULT_15	
ABB68417	
ID	ABB68417 standard; Protein; 743 AA.
XX	
AC	ABB68417;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 32043.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:35:15 ; Search time 28.21 Seconds
(without alignments)
289.528 Million cell updates/sec

Title: US-09-720-583a-3

Perfect score: 420
Sequence: 1 MTTRELRPRNGYSIAAANK.....EVGVSQIVHYALNNRIDA 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215	51.2	105	2 S32702	hypothetical prote
2	82.5	19.6	198	2 AC2478	transposase alr700
3	82	19.5	119	2 JS0053	hypothetical 13.7K
4	78	18.6	150	2 D90190	hypothetical prote
5	78	18.6	318	2 T35457	probable insertion
6	75.5	18.0	314	2 T31220	transposase homolo
7	74.5	17.7	217	2 A13429	transposase BME14
8	74.5	17.7	674	2 SS4178	low affinity penic
9	74	17.6	139	2 H75179	hypothetical prote
10	72.5	17.3	643	2 B75055	hypothetical prote
11	72.5	17.3	673	2 SS4175	low affinity penic
12	72.5	17.3	673	2 SS4182	low affinity penic
13	72.5	17.3	678	2 SS4176	low affinity penic
14	72.5	17.3	678	2 A36903	low affinity penic
15	72	17.1	88	2 T43610	probable IS1617 tr
16	72	17.1	113	2 T00224	hypothetical 12.7K
17	72	17.1	125	2 F87326	transposase
18	72	17.1	131	2 T36638	probable substrate
19	72	17.1	201	2 S38652	resolvasase - Klebsi
20	72	17.1	297	2 AF3651	transcription regu
21	71	16.9	151	2 H95394	probable transcrip
22	71	16.9	297	2 E95400	probable LysK-fam1
23	71	16.9	427	1 RRYC65	RNA-directed DNA p
24	70	16.7	211	2 F75593	DNA-binding respon
25	70	16.7	297	2 C98259	hypothetical prote
26	70	16.7	297	2 AF3025	transcription regu
27	70	16.7	340	2 T43559	probable transposa
28	70	16.7	340	2 AB0096	transposase for in
29	70	16.7	340	2 AB0198	transposase for in

30	70	16.7	340	2 AB0395	transposase for in
31	70	16.7	340	2 AB0488	transposase for in
32	70	16.7	340	2 AC0031	transposase for in
33	70	16.7	340	2 AC0247	transposase for in
34	70	16.7	340	2 AC0342	transposase for in
35	70	16.7	340	2 AD0070	transposase for in
36	70	16.7	340	2 AD0124	transposase for in
37	70	16.7	340	2 AD0139	transposase for in
38	70	16.7	340	2 AD0185	transposase for in
39	70	16.7	340	2 AD0206	transposase for in
40	70	16.7	340	2 AD0417	transposase for in
41	70	16.7	340	2 AD0457	transposase for in
42	70	16.7	340	2 AE0113	transposase for in
43	70	16.7	340	2 AE0163	transposase for in
44	70	16.7	340	2 AE0254	transposase for in
45	70	16.7	340	2 AE0292	transposase for in

ALIGNMENTS

RESULT 1
S32702
hypothetical protein 4 - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Jun-2000
C:Accession: S32702; S32200
R:Shi, S.L.; Wang, Z.X.; Deng, Y.; Zhu, H.; He, B.; Zheng, Z.X.
submitted to the EMBL Data Library, March 1993
A:Description: Complete nucleotide sequence of a plasmid pX210142 from Corynebacteriu
A:Reference number: S32699
A:Accession: S32702
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <SH1>
A:Cross-references: EMBL:X72691
C:Genetics:
A:Start codon: GTG
C:Superfamily: Corynebacterium glutamicum hypothetical protein 4

Query Match 51.2%; Score 215; DB 2; Length 105;
Best Local Similarity 53.3%; Pred. No. 7.8e-15;
Matches 40; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY 1 MTTRELRPRNGYSIAAANKIGVSESTVYKRWTSREPERFVARHARIRERSEGSQ 60
DB 19 MTKRTRIPRNGKTYIREVAEGTGLSTATIERWTSAPREDYLAQANEKRVQELRAKGLSM 78
QY 61 RAIAAEVGSVGVTH 75
DB 79 RAIAAEICGSVGLVH 93

RESULT 2
AC2478
transposase alr7003 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120alpha
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2478
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2478
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <KUR>
A:Cross-references: GB:BA000020; PTDN:BA078087.1; PTD:917135541; GSPDB:GN00180
C:Experimental source: strain PCC 7120
C:Genetics:

A:Accession:J311221
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule: DNA
A:Residues: 1-115,145-146, 'AMLSRTVP', 156, 'S' <R03>
A:Cross-references: EMBL:AF079317; NID:g3378281; PID:g3378381; PIDN:ADD03944.1
C:Genetics:
A:Gene: 18A
A:Genome: plasmid pNL1
A:Introns: 115/3

Query Match	18.0%	Score 75.5	DB 2	Length 314
Best Local Similarity	26.2%	Pred. No. 3.2		
Matches	28	Conservative	13	Mismatches 33; Gaps 4

```

QY 11 GYSIAAAKRLCVSESYTKRWITSEPRE--EVARVAARHARIRELRSEG----- 57
      | | | | : | | | : | | : | | :
Db 20 GLSCRAAAAREGVAPSTAIRWHAQRATGDFAPRPGGDMRSHRVEERSADILALWEARK 79

```

```
QY 58 -----QSMHAI AEVG--VSVGT VH-----YALNKNRTD 84
      : : | | | | | | : : : : |
Db 80 DISLEELRIALAEVGLTVSVAGLHREFVRRGMRGTRKKTGHAVEQDRPD 126
```

RESULT	7
AI3429	
transposase BMEI1423 [imported]	- <i>Brucella melitensis</i> (strain 16M)
C:Species: <i>Brucella melitensis</i>	

Ci.Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C.Accession: AI3429
R.DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muijer, C.; Los, R.
Mazur, M.; Colman, E.; Cohen, D.; Neufeld, J.; Oren, M.

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A1:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference number: AD3252; PMID:11756688
 A2:3252

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <CUR>
A:Accession: P0006017. EMDN: 1AF5C04.1. EMD: 17002423. Genes: 25002666

A: Experimental source: strain 16M
C: Genetics:
A: Gene: BME11423
A: Map: cont4199.t

Query Match	17.7%	Score	74.5;	DB 2;	Length	217;
Doc 1	20.2%					
Doc 2	20.2%					
Doc 3	20.2%					
Doc 4	20.2%					
Doc 5	20.2%					

Matches	28;	Conservative	7;	Mismatches	25;	Indels	27;	Gaps	3;
15	AAAKKTCGSESYVKRMTSEPRFEFVARVAARA	-----		RIRRL	53				

Db 51 AQAARLYGVSLKIVSRNT----ERF--RISGRAMTDRSSRPTRSRQMGDIVERILHL 104
Qy 54 RSEGOSMRAIAAEVGVSVGTIVHALNK 80

DB 105 RROQLTGKHIAMETGVPATVSRILRR 131
BESTUT 8

RESULT 8
S54178
low affinity penicillin-binding protein 5 (PBp5) - Enterococcus faecium (Fragment)
C:Species: Enterococcus faecium
C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 26-May-2000
Accession: S54178

A:Description: The low affinity penicillin-binding protein 5 in wild type and highly penicillin-resistant strains of *Staphylococcus aureus*

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-674 <ZOR>

A: Cross-references: EMBL:X84862; NID:g790432; PIDN:CAAS9289.1; PID:g790433
C: Superfamily: penicillin-binding protein 2B

Query Match	17.7%	Score 74.5;	DB 2;	Length 674;
Best Local Similarity	25.3%	Pred No. 8.8;		
Matches 19; Conservative	17;	Mismatches 24;	Indels 15;	Gaps 3

```
QY      8 PRNGSIALLA-----KLL--GVSESVYKRWTSEPREEVARVAARHARIRELRSGS 59
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB     366 PKTGDLALASPSPYDPNKMINGISOEDYKAYEEDPEQPISRATGYA-----PGST 418
```

QY 60 MRAIAAEVGVSVGTV 74
: | | : | : | :
Db 419 EKMITAIGLDNGTI 433

RESULT 9
H75179
hypothetical protein PAB2413 - Pyrococcus abyssi (strain Orsay)

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000
C;Accession: H75179
R;anonymous, Genoscope
Submitted to the Genbank Data Library Jul-2000

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: A75179
A:Status: complete genome

A: Molecule type: DNA
A: Residues: 1-139 <KAW>
A: Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49519.1; PID:g545190001; molref:1

C;Genetics:
A;Gene: PAB2413
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB2413

Query Match	17.6%	Score 74;	DB 2;	Length 139;
Best Local Similarity	23.3%	Pred. NO. 2;		
Matches 17; Conservative	20;	Mismatches 32;	Indels 4;	Gaps 1;

```
QY      8 PRNGISIAAAKKLLGSESTVCRWTSEREEFVAARVAARRIRELRSEGSMPATAAEV 67
        | : | ::| | : || : || :|| : || :
Db     63 PKVSKKVLRALEQMGLKIVPEVGGRGRKRYSYKL----MIEELKKCKTPREISRDL 118
```

QY 68 GVSVGTVHYALNK 80
| : : | : | |
Db 119 GIPRTLYYLKK 131

RESULT 10
B75055
hypothetical protein PAB1429 - *Pyrococcus abyssi* (strain Orsay)

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: B75055
R:anonymous, Genoscope

A:Description: *Pyrococcus abyssi* genome sequence: Insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: E75055

A: Molecule type: DNA
A: Residues: 1-643 <RAW>
A: Cross-references: GB:AJ248287; GB:AL096836; NID:g54558657; PIDN:CAB50335.1; PID:e451

A; Experimental
C; Genetics:
A; Gene: PAB1429

Query Match	17.38;	Score 72.5;	DB 2;	Length 643;
Best Local Similarity	29.5%;	Pred. No. 13;		

J. Bacteriol. 180, 5192-5202, 1998
A.Title: Structural organization of virulence-associated plasmids of *Yersinia pestis*.
A.Reference number: Z22578; MUID:98422474
A.Accession: T43610
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-88 <HUP>
A.Cross-references: EMBL:AF053946; NID:g2996222; PIDN:AAC6592.1; PID:g2996259
A.Experimental source: strain KIM
A.Pertty, R.D.; Straley, S.C.; Felherston, J.D.; Rose, D.J.; Gregor, J.; Blattner, F.R.
Infect. Immun. 66, 4611-4623, 1998
A.Title: DNA sequencing and analysis of the low-Ca²⁺-response plasmid pCD1 of *Yersinia* sp.
A.Reference number: Z22273; MUID:98427122
A.Accession: T42913
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-88 <PBR>
A.Cross-references: EMBL:AF074612; NID:g3822037; PIDN:AAC69821.1; PID:g3822101
A.Experimental source: strain KIM5
C.Genetics:
A.Genome: plasmid pCD1
A.Note: Y0084

	Query Match	17.1%	Score 72:	DB 2;	Length 88;
	Best Local Similarity	23.7%	Pred. No. 2;		
	Matches 14;	Conservative 16;	Mismatches 25;	Indels 4;	Gaps 1;
QY	11 GYSIAAAKKTGVSSTYKRWTSERREEVAVVAARHRIIRLRISGGSGMRALAAEVG	69			
	: : : :				
Db	21 GRVEVCHCKMIDISEFTYNN-----KKKGCGGVTELRRLRQLLEDENCRILKVLADLSL	75			

Search completed: October 4, 2002, 15:35:16
Job time: 128 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:36:31 ; Search time 15.85 Seconds

(Without alignments)
207,644 Million cell updates/sec

Title: US-09-720-583a-3

Perfect score: 420
Sequence: 1 MTRERLRPRNGYSIAAAAK.....EVGVSVGTVHYALNKNRTDA 85

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	73	17.4	112	1	INSN_SHIDY
2	71	16.9	427	1	RT65_MYXXA
3	69.5	16.5	221	1	Y013_TREPA
4	69	16.4	100	1	INN2_ECOLI
5	68	16.2	134	1	INN1_ECOLI
6	67.5	16.1	595	1	Y585_SCHPO
7	66.5	15.8	829	1	CADC_RABIT
8	66	15.7	363	1	ALPB_RABIT
9	66	15.7	363	1	ALPB_RAT
10	66	15.7	572	1	MOES_LYTV
11	65	15.5	200	1	YHUB_ECOLI
12	65	15.5	827	1	ATCU_RHIME
13	64.5	15.4	165	1	YGV6_YEAST
14	64	15.2	363	1	ALPB_SHEEP
15	64	15.2	827	1	ATCU_RHIME
16	63.5	15.1	326	1	YDJG_ECOLI
17	63.5	15.1	622	1	YAE7_YEAST
18	63	15.0	487	1	YOM5_CABEL
19	63	15.0	920	1	AKH1_MAIZE
20	62.5	14.9	196	1	RPC_BP163
21	62.5	14.9	800	1	PSCS_CABEL
22	62	14.8	148	1	RL19_HAUMA
23	62	14.8	196	1	UHPA_ECOLI
24	62	14.8	196	1	UHPA_SALTY
25	62	14.8	378	1	ILEU_PIG
26	62	14.8	2468	1	MAPE_HUMAN
27	61.5	14.6	219	1	PAA4_ECOLI
28	61.5	14.6	408	1	BNZD_PSEPU
29	61.5	14.6	409	1	TODA_PSEPU
30	61.5	14.6	452	1	MLTD_ECOLI
31	61.5	14.6	787	1	RELA_MYCLE
32	61.5	14.6	1127	1	LONH_PYRHO
33	61	14.5	88	1	LE25_LYCES

34	61	14.5	100	1	YIS1_SHISO	P16939 shigella so
35	61	14.5	301	1	YMB1_BACSU	P39592 bacillus su
36	61	14.5	308	1	LIPA_CHIMU	O9P12 chlamydia m
37	61	14.5	462	1	REGB_RHOSH	O5368 rhodobacter
38	61	14.5	848	1	MYSP_DIRIM	P13392 dirofilaria
39	61	14.5	879	1	MYSP_ONCVO	O02171 onchocerca
40	61	14.5	880	1	MYSP_BRUMA	O01202 brugia mala
41	60.5	14.4	232	1	YAGL_ECOLI	P77607 escherichia
42	60.5	14.4	298	1	VGL7_HSV11	O00116 ictaluriid h
43	60.5	14.4	312	1	YMB2_MYCTU	O50683 mycobacteri
44	60.5	14.4	414	1	YMA2_MYCTU	O10523 mycobacteri
45	60.5	14.4	745	1	HGL2_ARATH	P46607 arabidopsis

ALIGNMENTS

```
RESULT 1
ID INSN_SHIDY STANDARD; PRT; 112 AA.
AC P39213;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insn for insertion sequence element IS911.
GN INSN.
OS Shigella dysenteriae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=622;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11546;
RX MEDLINE=90299839; PubMed=2163395;
RA Prere M.F., Chandler M., Fayet O.;
RT "Transposition in Shigella dysenteriae: isolation and analysis of
  IS911, a new member of the IS3 group of insertion sequences.";
RL J. Bacteriol. 172:4090-4099(1990).
CC -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
  SEQUENCE IS911.
CC -1- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 8.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17613; -; NOT ANNOTATED CDS.
DR InterPro; IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
KW Transposable element; Transposition; DNA-binding; DNA recombination.
SQ
SEQUENCE 112 AA; 12739 MW; 7B01B52B5E6F86A CRC64;
```

Query Match 17.4%; Score 73; DB 1; Length 112;
Best Local Similarity 34.0%; Pred. No. 0.74;
Matches 18; Conservative 9; Mismatches 20; Indels 6; Gaps 1;

12 YSIAAARKKGVSESTVKTSEPREFEVAR-----VAARHARIRELRSCQ 58
11:|||||:|||||:|||||:|||||:|
35 YTVADAAKADVGLSTMTRWVKQJLRDROGCTPKASPTTPQIIRLRKKLQ 87

RESULT 2
ID RT65_MYXXA STANDARD; PRT; 427 AA.
AC P23071;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

DE RNA-directed DNA polymerase from retron MX65 (EC 2.7.7.49) (Reverse
DE transcriptase) (MX65-RT).
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90138993; PubMed=1689062;
RA Inouye S., Herzer P.J., Inouye M.;
RT "Two independent retrons with highly diverse reverse transcriptases
RT in Myxococcus xanthus";
RL Proc. Natl. Acad. Sci. U.S.A. 87:942-945(1990).
CC -1- FUNCTION: PARTICIPATES IN THE SYNTHESIS OF A MSDNA (A BRANCHED RNA
CC LINKED BY A 2',5'-PHOSPHODIESTER LINKAGE TO A SINGLE-STRANDED DNA)
CC OF UNKNOWN FUNCTION. THE RETRON TRANSCRIPT SERVES AS PRIMER AND
CC TEMPLATE FOR THE REACTION, AND CODES FOR THE RT.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + [DNA](N).
CC -1- MISCELLANEOUS: M.XANTHUS CONTAINS TWO INDEPENDENT AND UNLINKED
CC RETRONS: MX65 AND MX162.
CC -1- MISCELLANEOUS: RETRONS MAY BE THE ANCESTORS OF RETROVIRUS.
CC -1- SIMILARITY: TO OTHER REVERSE TRANSCRIPTASES IN BACTERIA AND
CC RETROVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M30609; AAA8323.1; -.
DR PIR; A34864; RRYC65.
DR InterPro: IPR000123; RNA_DNAPolym.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt; 1.
DR PRINTS; PR00866; RNDNAPOLMS.
DR Transposase; RNA-directed RNA polymerase; Transposable element.
SQ SEQUENCE 427 AA; 48024 MW; C96436BF864D0D7 CRC64;

```

```

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE001186; AAC65013.1; -.
DR TIGR; TP0013; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 221 AA; 23872 MW; 5E1B311377B49DB2 CRC64;

```

Query Match 16.5%; Score 69.5; DB 1; Length 221;
Best Local Similarity 32.2%; Pred. No. 3.4;
Matches 19; Conservative 7; Mismatches 20; Indels 13; Gaps 1;

```

OY 11 GYSIAAAKKILGVSESTVYKRWTSREDFEVARVAARHARIRELRSEGSMAIAEVCV 69
DB 161 GYAIWMAKKKMA-----RTLVHVSFPRHLSKALERTROYMEAVCADIGI 206

```

RESULT 4
ID INN2_ECOLI STANDARD; PRT; 100 AA.
AC P39212;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insn for insertion sequence element IS911B.
GN INSN2 OR B4283.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
CC SEQUENCE IS911.
CC -1- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U14003; AA97179.1; -.
DR EMBL; AE000499; AAC77239.1; -.
DR EcoGene; EG40014; insn.
DR InterPro: IPR002514; Transposase_8.
DR Pfam: PF01527; Transposase_8; 1.
KW Transposable element; Transposition; DNA-binding; DNA recombination;

OY 61 RAIA-----AEVGVSVGIV 74
 Db 116 RIIVAGEGKERYGHAFYADSGNSVGEI 142

RESULT 7
 CADG_RABIT STANDARD; PRT; 829 AA.

AC Q28634;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
 GN COH16.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A. AND OF SEQUENCE 570-586.

RC STRAIN-NEW ZEALAND WHITE;
 RX MEDLINE=95340560; PubMed=7615566;
 RA Thomson R.B., Igarashi P., Biemesderfer D., Kim R., Abu-Alfa A.,
 RA Soleimani M., Aronson P.S.;
 RT "Isolation and cDNA cloning of ksp-cadherin, a novel kidney-specific
 RT member of the cadherin multigene family."
 RL J. Biol. Chem. 270:17594-17601(1995).

CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: KIDNEY-SPECIFIC. LIMITED TO THE BASOLATERAL
 CC MEMBRANES OF RENAL TUBULAR EPITHELIAL CELLS.
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC EMBL: U28945; AAC48472.1; -.
 DR HSSP: P09803; 150H.
 DR InterPro: IPR002126; Cadherin.
 DR Pfam: PF00028; cadherin; 6.
 DR SMART: SM00112; CA: 6.
 DR PROSITE: PS00232; CADHERIN_1; 2.
 DR PROSITE: PS50268; CADHERIN_2; 6.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 829 POTENTIAL.
 FT DOMAIN 19 786 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 787 807 POTENTIAL.
 FT DOMAIN 808 829 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 25 126 CADHERIN 1.
 FT DOMAIN 131 235 CADHERIN 2.
 FT DOMAIN 242 336 CADHERIN 3.
 FT DOMAIN 341 449 CADHERIN 4.
 FT DOMAIN 455 564 CADHERIN 5.
 FT DOMAIN 569 665 CADHERIN 6.
 FT DOMAIN 666 786 ECTODOMAIN 6.
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 829 AA; 88827 MM; D2DF10E6C47A43B9 CRC64;

Query Match 15.8%; Score 66.5; DB 1; Length 829;

Best Local Similarity 26.8%; Pred. No. 26;
 Matches 22; Conservative 12; Mismatches 17; Indels 31; Gaps 4;

OY 31 WTSEPREFEVAVARAHRIET-----RSEGSMAIA-----AEVGVSVG 72
 Db 734 WV-EPREHIVPVVVSQARVWOLPVRVYVCCNTEGECMEKRVGKMPKLSAVGILVG 792

OY 73 TV-----HYALNKR 82
 Db 793 TLAIGFFLIIFTHLARKK 814

RESULT 8
 ALFB_RABIT STANDARD; PRT; 363 AA.

AC P79226;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).
 GN ALDOB OR ALDB.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Berardini T.Z., Amsden A.B., Penhoet E.E., Tolian D.R.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glyceraldehyde 3-phosphate + D-glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.

CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- MISCELLANEOUS: IN VERTEBRATES, THREE FORMS OF THIS UBQUITOUS
 CC GLYCOLYTIC ENZYME ARE FOUND, ALDOLASE A IN MUSCLE, ALDOLASE B IN
 CC LIVER, & ALDOLASE C IN BRAIN.
 CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE
 CC FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC EMBL: U85645; AAB42087.1; -.
 DR HSSP: P00883; IADO.
 DR InterPro: IPR000741; Aldolase_I.
 DR Pfam: PF00274; glycolytic-enz; 1.
 DR ProDom: PD001128; Aldolase_I; 1.
 DR PROSITE: PS00158; ALDOLASE_CLASS_I; 1.
 KW Lyase; Schiff base; Glycolysis; Multigene family.
 FT INIT MET 0
 FT BINDING 55 55 BY SIMILARITY.
 FT BINDING 146 146 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
 FT BINDING 229 229 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
 FT ACT_SITE 363 363 SCHIFF-BASE WITH DIHYDROXYACETONE-P.
 FT ESSENTIAL FOR ENHANCED ACTIVITY OF THE
 FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
 FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
 SQ SEQUENCE 363 AA; 39474 MM; 0A7185A72E89F436 CRC64;

Query Match 15.7%; Score 66; DB 1; Length 363;
 Best Local Similarity 25.6%; Pred. No. 13;
 Matches 22; Conservative 14; Mismatches 32; Indels 18; Gaps 2;

OY 5 ERLPRNGISIAAARKLGSESTVAKR-----TSERPEE-----VARVAR 46
 Db 20 ORIVANGKIIILADESVGTGMRLORIKRYENTENRQREILFTVDNSINOSIGVILF 79
 OY 47 HARIRELRSEGSMAIAAEVGVSVG 72

Db 80 HETLYKDSQKLEFRNLKEKGIYVG 105

RESULT 9

ALFB_RAT STANDARD; PRT; 363 AA.

AC P00884; P70706; 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DE 15-DEC-1998 (Rel. 37, Last annotation update)

GN Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_Taxid=10116;

RP SEQUENCE FROM N.A.

RX MEDLINE=85054935; PubMed=6094564; Tsutsumi K., Mukai T., Tsutsumi R., Mori M., Dalmon M., Tanaka T., Yatsuki H., Hori K., Ishikawa K.; "Nucleotide sequence of rat liver aldolase B messenger RNA." J. Biol. Chem. 259:14572-14575(1984).

RT [2]

RL SEQUENCE FROM N.A.

RN STRAIN=SPRAGUE-DAWLEY; MEDLINE=85160837; PubMed=2580098; Tsutsumi K., Mukai T., Tsutsumi R., Hidaka S., Arai Y., Hori K., Ishikawa K.; "Structure and genomic organization of the rat aldolase B gene." J. Mol. Biol. 181:153-160(1985).

RA [3]

RN SEQUENCE OF 184-363 FROM N.A.

RX MEDLINE=83213364; PubMed=6304044; Tsutsumi K., Mukai T., Hidaka S., Miyahara H., Tsutsumi R., Tanaka T., Hori K., Ishikawa K.; "Rat aldolase isozyme gene." J. Biol. Chem. 258:6537-6542(1983).

RL -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glyceraldehyde + D-glyceraldehyde 3-phosphate.

CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.

CC -1- SUBUNIT: HOMOTETRAMER.

CC -1- MISCELLANEOUS: IN VERTEBRATES, THREE FORMS OF THIS UBQUITOUS GLYCOLYTIC ENZYME ARE FOUND, ALDOLASE A IN MUSCLE, ALDOLASE B IN LIVER, & ALDOLASE C IN BRAIN.

CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M10149; AAA40716.1; -

CC EMBL; X02284; CAA26156.1; -

CC EMBL; X02285; CAA26156.1; JOINED.

CC EMBL; X02286; CAA26156.1; JOINED.

CC EMBL; X02287; CAA26156.1; JOINED.

CC EMBL; X02288; CAA26156.1; JOINED.

CC EMBL; X02289; CAA26156.1; JOINED.

CC EMBL; X02290; CAA26156.1; JOINED.

CC EMBL; X02291; CAA26156.1; JOINED.

CC EMBL; Y01223; CAA24533.1; -

CC PIR; A22585; ADRTB.

CC HSSP; P00883; IADO.

CC InterPro; IPR000741; Aldolase_I.

CC Pfam; PF00274; glycolytic_enzy; 1.

CC ProDom; PD001128; Aldolase_I; 1.

CC PROSITE; PS00158; ALDOLASE_CLASS_I; 1.

KM Lyase; Schiff base; Glycolysis; Multigene family.

FT INT_MET 0

FT BINDING 55

FT BINDING 146

FT BINDING 229

FT ACT_SITE 363

FT

FT CONFLICT 233

FT SEQUENCE 363 AA; 39487 MW; 8E4F3E621E6DD9A CRC64;

Query Match 15.7% Score 66; DB 1; Length 363;

Best Local Similarity 25.6% Pred. No. 13;

Matches 22; Conservative 14; Mismatches 32; Indels 18; Gaps 2;

QY 5 ERLPRNGYSIAAKKIGVSESTVGRW---TSPRREF-----VARPAAR 46

DB 20 QRIANGKGIADSESYGMNRLQRIKVENTENRRQFRELFSVDSNISQISGVILF 79

QY 47 HARIRELRSEGSRAIAEYGSVG 72

DB 80 HETLYKDSQKLEFRNLKEKGIYVG 105

RESULT 10

MOES_LYTVA STANDARD; PRT; 572 AA.

ID MOES_LYTVA

AC P52962; 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Moesin.

OS Lytechinus variegatus (Sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinzoa; Echinozoa; Euechinozoa; Echinoidae; Echinacea; Temnopneuroidea; Toxopneustidae; Lytechinus.

OX NCBI_Taxid=7654;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95256329; PubMed=7738093; Bachman E.S., McClay D.R.; "Characterization of moesin in the sea urchin Lytechinus variegatus: RT distribution to the plasma membrane following fertilization is inhibited by cytochalasin B." J. Cell Sci. 108:161-171(1995).

RL -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL STRUCTURES TO THE PLASMA MEMBRANE.

CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U14180; AAC46514.1; -

CC InterPro; IPR000299; Band_4.1.

CC InterPro; IPR000798; Ezrin_radixin_moesin.

CC Pfam; PF00373; Band_41; 1.

CC Pfam; PF00769; ERM; 1.

CC PRINTS; PR00935; BAND41.

CC SMART; SM00295; B41; 1.

CC PROSITE; PS00660; BAND_41_1; 1.

CC PROSITE; PS00661; BAND_41_2; 1.

CC PROSITE; PS50057; BAND_41_3; 1.

CC Structural protein; Cytoskeleton.

FT DOMAIN 58 224 BAND 4.1-LIKE.

SO SEQUENCE 572 AA; 67579 MW; E1BA3F8B845E764 CRC64;

GN ACTP2 OR ATC2 OR RB1018 OR SMB21578.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE-21396508; PubMed-11481431;
RA Flann T.M., Meldner S., Wong K., Bunrmester J., Chain P.,
RA Vorholster F.J., Hernandez-Lucas I., Becker A., Cowie A., Guzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: INVOLVED IN COPPER TRANSPORT (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC -1- SIMILARITY: CONTAINS 2 HMA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL603645; CAC9418.1; -;
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
DR PROSITE; PS01047; HMA_1; 2.
DR PROSITE; PS50846; HMA_2; 2.
KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW Metal-binding; Copper; Repeat; Plasmid; Complete proteome.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 458 478 POTENTIAL.
FT TRANSMEM 771 793 POTENTIAL.
FT TRANSMEM 797 819 POTENTIAL.
FT DOMAIN 16 81 HMA 1.
FT DOMAIN 83 149 HMA 2.
FT METAL 26 26 COPPER (POTENTIAL).
FT METAL 29 29 COPPER (POTENTIAL).
FT METAL 93 93 COPPER (POTENTIAL).
FT METAL 96 96 COPPER (POTENTIAL).
FT MOD_RES 515 515 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 714 714 MAGNESIUM (BY SIMILARITY).
FT METAL 718 718 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 827 AA; 85861 MW; A3D8DFDD1315FCB CRC64;

Query Match 15.2%; Score 64; DB 1; Length 827;
Best Local Similarity 29.9%; Pred. No. 47;
Matches 23; Conservative 11; Mismatches 17; Indels 26; Gaps 3;

QY 10 NGYSTAAAKKLGVSSTVKTSTSEPREFVARV--AARHARIRLRSEGSRAIA--- 64
Db 670 NRRFAEAIARKLGI-----DEVAEVLPEKGVKRLRKGSRVAFITGDI 716
QY 65 -----AEVCVSYGT 73
Db 717 NDAPALAEADVGIAGVT 733

THIS PAGE BLANK (USPTO)